Law and Human Behavior, Vol. 25, No. 5, October 2001 (© 2001)

Asking the Gatekeepers: A National Survey of Judges on Judging Expert Evidence in a Post-*Daubert* World[1]

Sophia I. Gatowski,[2,6] Shirley A. Dobbin,[2] James T. Richardson,[3] Gerald P. Ginsburg,[4] Mara L. Merlino,[5] and Veronica Dahir[5]

Drawing on the responses provided by a survey of state court judges (N = 400)*, empirical evidence is presented with respect to judges' opinions about the* Daubert *criteria, their utility as decision-making guidelines, the level to which judges understand their scientific meaning, and how they might apply them when evaluating the admissibility of expert evidence. Proportionate stratified random sampling was used to obtain a representative sample of state court judges. Part I of the survey was a structured telephone interview (response rate of 71%) and in Part II, respondents had an option of completing the survey by telephone or receiving a questionnaire in the mail (response rate of 81%). Survey results demonstrate that judges overwhelmingly support the "gatekeeping" role as defined by* Daubert*, irrespective of the admissibility standard followed in their state. However, many of the judges surveyed lacked the scientific literacy seemingly necessitated by* Daubert*. Judges had the most difficulty operationalizing falsifiability and error rate, with only 5% of the respondents demonstrating a clear understanding of falsifiability and only 4% demonstrating a clear understanding of error rate. Although there was little consensus about the relative importance of the guidelines, judges attributed more weight to general acceptance as an admissibility criterion. Although most judges agreed that a distinction could be made between "scientific" and "technical or otherwise specialized" knowledge, the ability to apply the* Daubert *guidelines appeared to have little bearing on whether specific types of expert evidence*

[1]Earlier versions of this paper were presented at the meeting of the American Psychology-Law Society, New Orleans, LA, March 2000, the meeting of the American Psychology-Law Society and the European Society and Law Association, Dublin, Ireland, July 1999, the meeting of the Law and Society Association, Aspen, CO, May 1998, and the National Conference on Science and the Law, National Institute of Justice and the National Academies of Science, San Diego, CA, October 2000.

[2]National Council of Juvenile and Family Court Judges and Grant Sawyer Center for Justice Studies, University of Nevada, Reno, Nevada.

[3]Judicial Studies Program and Department of Sociology, University of Nevada, Reno, Nevada.

[4]Department of Psychology, University of Nevada, Reno, Nevada.

[5]Interdisciplinary PhD Program in Social Psychology and Grant Sawyer Center for Justice Studies, University of Nevada, Reno, Nevada.

[6]To whom correspondence should be addressed at Permanency Planning for Children Department, National Council of Juvenile and Family Court Judges, P.O. Box 8970, University of Nevada, Reno, Nevada 89507; e-mail: gatowski@pppncjfcj.org.

were designated as "science" or "nonscience." Moreover, judges' "bench philosophy of science" seemed to reflect the rhetoric, rather than the substance, of Daubert. *Implications of these results for the evolving relationship between science and law and the ongoing debates about* Frye, Daubert, Joiner, *and* Kumho *are discussed.*

In the wake of *Daubert* (*Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 1993) and subsequent cases upholding and extending its scope and applicability, a large body of scholarship continues to debate the merits of the *Daubert* criteria as judicial decision-making guidelines. Included in this debate are discussions of the relative importance of each of the criteria to the admissibility decision and procedures for their application (e.g., Berger, 1994; Dwyer, 1994; Faigman, Kaye, Saks, & Sanders, 1997, 1999; Fenner, 1996; Krauss & Sales, 1999; Richardson, 1994; Schwartz, 1997; Watkins, 1994), the extent to which judges understand and can properly apply the criteria when assessing the validity and reliability of proffered scientific evidence (e.g., Beecher-Monas, 1999; Faigman et al., 1997, 1999; Grove & Barden, 1999; Krauss & Sales, 1999; Richardson, Dobbin, Gatowski, Ginsburg, Merlino, & Dahir, 1998; Richardson, Ginsburg, Gatowski, & Dobbin, 1995; Saxe & Ben-Shakar, 1999; Schwartz, 2000), the potential differential application of the criteria to various domains of expert testimony and the implications of their application for the admissibility and legitimacy of different domains of knowledge (e.g., Agrimonti, 1995; de Vyver, 1999; Faigman, 1995; Faigman et al., 1997, 1999; Faigman & Wright, 1997; Goodman-Delahunty, 1997; Gottesman, 1994; Grove & Barden, 1999; Krauss & Sales, 1999; Laser, 1997; Needham, 1998; Odgers & Richardson, 1995; Penrod, Fulero, & Cutler, 1995; Renaker, 1996; Richardson, Dobbin, Gatowski, Ginsburg, Merlino, Dahir, & Cotton 1998; Richardson, Ginsburg, et al., 1995; Saxe & Ben-Shakar, 1999; Showalter, 1995; Shuman & Sales, 1999; Slobogin, 1998, 1999). Although almost all of this past discourse about *Daubert* and its progeny recognizes that judges are central and active figures in admissibility decision-making, and becoming increasingly more so as a result of recent legal decisions, rarely have judges themselves been asked for their opinions and experiences with respect to *Daubert*, their gatekeeping role, and the admissibility decision-making process.

Typically, empirical research about the impact of *Daubert* has focused on analysis of published appellate opinions (cf. Groscup, Studebaker, Huss, O'Neil, & Penrod, 2000; Richardson, Dobbin, Gatowski, Ginsburg, Merlino, Dahir, & Cotton, 1998), often imposing artificial criteria (e.g., number of words devoted to discussion of specific *Daubert* guidelines or issues) to infer conclusions about the utility and relevance of *Daubert* to admissibility decisions, the scientific literacy of judges, and the extent to which there is a differential application of *Daubert* to different domains of expert testimony. While providing important insight regarding the influence of *Daubert*, an empirical analysis of published case law is, by its very nature, restricted to an analysis of post hoc justifications of those writing a decision in a particular case and does not fully capture the judicial decision-making process. Although an empirical analysis of case law provides important data about judges' normative, case specific reasoning, research has demonstrated that there may be significant differences

between published and unpublished cases, and that these differences may be dependent upon the case characteristics analyzed and the legal questions involved (Siegelman & Donohue, 1990; Songer, 1988). Thus, information about judges' decision-making processes obtained from judicial opinions may differ in important ways from information obtained when judges are simply asked to talk about the process of making admissibility decisions in a survey or interview format. Therefore, caution must be exercised when generalizing from the population of published cases to judges' actual decision-making practices as a whole. Although judges who are being surveyed or interviewed are still speaking from their position as a judge, they may be less constrained because their anonymity is protected and the potential for appellate review is removed. Moreover, in a survey or interview format judges may not necessarily be asked to tie their responses to specific cases, but rather to base their answers on their general experiences with and opinions about expert evidence admissibility issues. These features of the survey process should permit judges' opinions of *Daubert* and their understanding of its application to emerge less encumbered by measured reasoning in relation to a specific case.

To add to what is being learned from analyses of case law and in order to shed light on the judicial perspective with respect to current debates about *Daubert*, we turned to the gatekeepers themselves with the following overarching questions in mind.

- What do judges think about the intent and value of *Daubert*?
- Do judges see the role of "gatekeeper" as an appropriate one?
- Can judges operationalize *Daubert*'s scientific concepts and appropriately use them as decision-making guides?
- Do judges attribute equal importance to each of the guidelines?
- Do judges apply the guidelines differently to different domains of knowledge?
- Does the judicial "bench philosophy of science" reflect the philosophical assumptions of *Daubert*?

Drawing on the responses provided by a national survey of state trial court judges ($N = 400$), this paper presents empirical evidence regarding judicial opinions about the *Daubert* criteria, their utility as decision-making guidelines, and their applicability to different domains of expert knowledge. By asking judges directly, the results of this national survey provide important information about whether judges are able to operationalize the *Daubert* criteria; whether some *Daubert* criteria are of more relative importance in the decision-making process than others; and what types of expert knowledge judges define as "scientific" and in what ways, if any, the *Daubert* guidelines help judges to make such determinations. It is important to note that the survey was completed prior to the passage of *Kumho Tire Co. v. Patrick Carmichael* (1999) that clarified that the *Daubert* analysis applies to scientific, technical and otherwise specialized knowledge, and not exclusively to scientific knowledge (Kumho, at 1175). The implication of the survey results, in light of *Kumho*, will be addressed in the conclusion of this article.

The Legal Landscape: Frye, Daubert, and Beyond. In 1923, the United States Court of Appeals for the D.C. Circuit held in *Frye v. United States* (1923) that to

be admissible, proffered testimony by a scientific expert must be based on a discovery or principle that has "gained general acceptance in the particular field in which it belongs" (*Frye*, at 1014). At the core of *Frye* was the contention that judges should defer to scientists, that is, acceptance by scientists was the sole criterion for determining whether a particular domain of knowledge constituted a genuine area of scientific expertise about which a properly qualified expert could testify. *Frye*'s "general acceptance" rule became the rule in both federal and state courts.

In contrast to *Frye*, in *Daubert v. Merrell Dow Pharmaceuticals, Inc.* (1993) the Court held that testimony will be classified as scientific, and thus presented to a jury as expert testimony, only if a judge first determines that the proffered testimony consists of inferences and assertions "derived by the scientific method" (*Daubert*, at 590). In *Daubert*, the Court explicitly placed judges in the role of "gatekeepers" who evaluate the scientific validity and reliability of scientific evidence (*Daubert*, at 597). *Daubert*'s central premise is that judges can, and must, decide whether proffered scientific testimony is based on the scientific method without taking a position regarding the veracity of particular scientific conclusions. Judges are advised that while deciding whether to admit the scientific evidence, "[t]he focus . . . must be solely on principles and methodology; not on the conclusions they generate" (*Daubert*, at 595).

In *Daubert*, the Supreme Court focused on Federal Rule of Evidence (FRE)[7] 702 in order to emphasize that the "subject of an expert's testimony must be 'scientific . . . knowledge.' " The Court explained that ' "scientific' implies a grounding in the methods and procedures of science," whereas ' "knowledge' connotes more than subjective belief or unsupported speculation" (*Daubert*, at 595). Thus, "in order to qualify as 'scientific knowledge,' an inference or assertion must be derived by the scientific method" and must be "supported by appropriate validation" (*Daubert*, at 595). Discussing the nature of scientific knowledge, the Court quoted both Webster's dictionary and two amicus briefs (one from a group of scientists, the other jointly authored by the American Association for the Advancement of Science and the National Academy of Science) to caution that "[s]cience is not an encyclopedic body of knowledge about the universe. Instead it represents a process for proposing and refining theoretical explanations about the world that are subject to further testing and refinement" (*Daubert*, at 590).

The Court in *Daubert* does not explicitly define science, nor does the court "presume to set out a definitive checklist or test" (*Daubert*, at 593). However, the Court does offer "some general observations" (*Daubert*, at 593) that trial judges should bear in mind when they assess expert evidence. Instead of looking to general acceptance as the sole standard, as was the case under *Frye*, the trial judge post-*Daubert* must assess

[7]FRE 702: "If scientific, technical, or other specialized knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue, a witness qualified as an expert by knowledge, skill, experience, training, or education, may testify thereto in the form of an opinion or otherwise." A witness may, therefore, testify as an expert when three conditions are met: (1) there is distinctive knowledge, (2) which will aid the jury, and (3) the witness is a qualified expert. Note, the Proposed Amendment to Rule 702 states: "If scientific, technical, or otherwise specialized knowledge . . . the form of an opinion or otherwise: *provided that (1) the testimony is sufficiently based upon reliable facts or data, (2) the testimony is the product of reliable principles and methods, and (3) the witness has applied the principles and methods reliably to the facts of the case.*"

the scientific merits of proffered testimony with respect to four general guidelines. The Court recognized as "guidelines" for judicial consideration: (1) "falsification" or whether a theory or technique "can be (and has been) tested" (*Daubert*, at 593); (2) the "known or potential rate of error" associated with a "particular scientific technique" and the "existence and maintenance of standards controlling the technique's operation" (*Daubert*, at 594); (3) whether the theory or technique has been the subject of "peer review and publication" (*Daubert*, at 593); and (4) the "general acceptance" of the proposed testimony in the scientific community (*Daubert*, at 594).

It is important to note that the *Daubert* majority explicitly declined to decide whether its four factors were either necessary or sufficient components for an adequate assessment of the scientific method. The Court notes that "the inquiry envisioned by Rule 702 is . . . a flexible one" (*Daubert*, at 594). However, while giving the judiciary discretion in applying the *Daubert* guidelines in the performance of their gatekeeping role, the Court provided little if any guidance as to the meaning or application of the guidelines. For example, the Supreme Court clearly emphasized falsifiability as the cornerstone of the scientific method (*Daubert*, at 593), yet the Court neither explained nor gave examples of how to use falsifiability as a decision-making guideline. The Court also provided minimal guidance as to how a judge is to determine or evaluate the error rate of a particular scientific technique, or how to evaluate controls and possible confounding factors. Nor did the Court acknowledge that different scientific disciplines may have different ways of controlling and estimating errors, and how this is to be considered in a *Daubert* analysis of proffered expert evidence.

The Court's expressed intention in *Daubert* was to switch the trial courts' focus from whether the expert's conclusions had garnered a scientific consensus to whether the expert's techniques and methodology were valid (*Daubert*, at 595). That does not mean, as the Court later explained in *General Electric Co. v. Joiner* (1997), that the trial judge could ignore the expert's conclusions. Rather, the trial court must examine the expert's techniques and methodologies for consistency with the expert's conclusions and with the facts of the case at hand (the question of "fit"). Conclusions and methodology, as the Court pointed out in *Joiner*, are not entirely distinct from one another, and there must be a valid connection between them (*Joiner*, at 146).

The *Daubert* decision also did not provide clear guidance as to whether the four guidelines articulated should apply only to "scientific" knowledge or to "scientific" knowledge and "technical or otherwise specialized" knowledge. And, if such a distinction between types of knowledge is to be made, *Daubert* did not provide any guidance as to how such a distinction should be made. Since the *Daubert* decision, courts have applied *Daubert*'s "scientific validity" test to a wide variety of expert testimony, but there has been considerable variability with respect to whether fields such as psychology and other so-called "soft sciences" constitute science and whether and how the *Daubert* factors should be applied to nonscientific knowledge (cf. *Hawthorne Partners v. AT&T Technologies., Inc.*, 1993; *Iacobelli Construction, Inc., v. County of Monroe*, 1994; *In re Aluminum Phosphide Antitrust Litigation*, 1995; *Isely v. Capuchin Province*, 1995; *Moore v. Ashland Chemical, Inc.*, 1997; *Petruzzi's IGA Supermarkets, Inc., v. Darling-Delaware Co.*, 1993; *Williams v. General Motors Corporation*, 1994. See also Agrimonti, 1995; Faigman et al., 1997, 1999; Goodman-Delahunty, 1997;

Groscup et al., 2000; Grove & Barden, 1999; Needham, 1998; Penrod et al., 1995; Renaker, 1996; Richardson, Dobbin, Gatowski, Ginsburg, Merlino, Dahir, & Colton, 1998; Richardson, Ginsburg, et al., 1995; Risinger & Saks, 1996; Showalter, 1995; Shuman & Sales, 1999; Slobogin, 1998, 1999).

Whether or not the *Daubert* guidelines apply to all forms of technical or otherwise specialized knowledge, or just scientific knowledge, was addressed by the United States Supreme Court in *Kumho Tire Co. v. Patrick Carmichael* (1999). Reversing the 11th Circuit, the Court held in *Kumho* that the factors for a court to use in determining the reliability of a scientific theory or technique, as set out in *Daubert*, may apply to testimony of engineers (the expert evidence under consideration in *Kumho*) and other experts who are not scientists. The Court noted that *Daubert* set forth a trial judge's "gatekeeping" obligation under FRE 702 to ensure that expert testimony is relevant and based on reliable scientific theories and that FRE 702 applies to all expert testimony because the language of FRE 702 does not distinguish between "scientific," "technical," or "other specialized" knowledge. Justice Breyer concluded that the line between scientific and nonscientific evidence is unclear and "conceptual efforts to distinguish between the two are unlikely to produce clear legal lines capable of application in particular cases" (*Kumho*, at 1175). In *Kumho*, the Court further strengthened the discretion of the trial judges' gatekeeping role in noting that the judge had "broad latitude" and "considerable leeway" (*Kumho*, at 1176) in deciding how to assess the validity of different forms of nonscientific expert knowledge.

METHOD

The primary purpose of the national survey was to assess the level to which the judiciary understand the scientific meaning of the *Daubert* guidelines and how they might apply them when evaluating the admissibility of scientific evidence. In addition to assessing the scientific literacy of judges, the survey also asked respondents for their opinions about the relevance and utility of the *Daubert* criteria to the judicial gatekeeping role and admissibility decision-making process.

Survey Sample

Construction of the sampling frame began with the generation of a list of eligible sample elements using the 1997–1998 edition of *The American Bench* (1997/1998) a biennial publication containing information about the structure of federal and state court systems, as well as rosters of all the judges by state (including names, addresses, telephone and fax numbers), organized by the level of jurisdiction within each state court system. To be included in the sampling frame, a judge either had to be sitting on the bench of the state trial court of general jurisdiction or on the bench of a court of special jurisdiction hearing a docket likely to contain the types of evidence of interest to this research. Because some states listed vacant judicial seats at the time of publication, the initial sampling frame generated from *The American Bench* was supplemented as necessary using current rosters from the Administrative Office of

the Courts for those states (see Dobbin et al., 2001 more detail). The final sampling frame consisted of 9,715 state trial court judges from all 50 states and the District of Columbia.

Judges in the sampling frame were first stratified by federal circuit and then by state. Proportionate stratified random sampling with a constant sampling fraction was used to obtain a sample that was both representative of geographical distribution of judges and levels of court. Given the complexity and length of the telephone survey, and the status and professional distance of the sample population, an estimated sample size of 1,264 was drawn in order to achieve the goal of 400 completed interviews (Lavrakas, 1993). This sample of 1,264 judges was then divided into two replicates, using an odd–even split (i.e., judges number 1,3,5, etc. in Sample Replicate A ($n_A = 643$); judges number 2,4,6, etc. in Sample Replicate B ($n_B = 621$)). All judges in Sample Replicate A were contacted first. Those in Sample Replicate B were held in reserve and were to be surveyed only if the goal of 400 completed surveys was not met using Sample Replicate A. Sample Replicate B was not used as sampling from Replicate A proved sufficient (see Dobbin et al., 2001 for more detail).

Survey Instrument

After several months of development based on a substantive literature review and case law analysis, a draft survey instrument was reviewed by members of the research project's national advisory committee and pretested on focus groups of judges attending classes at the National Judicial College and the Judicial Studies Program at the University of Nevada, Reno. Judges in the pretest groups reacted negatively to the initial survey instrument, commenting that they were being "tested" on scientific terms and methods and that this was not appropriate given their professional role as judges. When research staff reviewed the pretest responses, however, they found that the questions were addressing the appropriate issues and eliciting topic-specific responses. Nevertheless, modifications were made to the wording of questions and to the general structure of the survey instrument to be sensitive to expressed concerns about "testing" judges' knowledge of science. For example, rather than asking for definitions of concepts such as falsifiability, survey questions asked judges to discuss how they would use the concept of falsifiability when making admissibility decisions (i.e., "How would you use the guideline of falsifiability when scientific expert testimony is proffered in your court?"). Their level of understanding of the criterion was then inferred from their responses using a standardized coding procedure (see discussion under coding and analysis of survey responses). These pretests proved critical to the final development of the survey instrument, not only ensuring that the survey was tapping the appropriate domain of knowledge, but also ensuring that the structure and wording of questions reflected a sophisticated understanding of the judicial role in admissibility decision-making.

The final survey instrument consisted of two parts and combined both telephone and mail methods. Part I of the survey was a structured telephone interview that focused on six major topics: (1) the standard of admissibility used in the respondents'

state; (2) the utility and application of the admissibility criteria specifically outlined in *Daubert*; (3) general opinions on issues surrounding the *Daubert* decision (e.g., judgments as to the decision's value and intent, appropriateness of the gatekeeper role, potential differential application of criteria to different domains of knowledge); (4) opinions on general aspects of the legal system (e.g., judgments regarding the over- or underreliance on experts in the courtroom and whether jurors can understand complex scientific evidence); (5) judicial designation of expert testimony as "scientific" or as "technical or otherwise specialized knowledge"; and (6) respondents' educational background and other demographic characteristics. The telephone interview in Part I was the more complex part of the survey involving, for example, multiple skip patterns and a significant number of open-ended questions.

Questions on Part II were directed toward judges' level of experience with specific types of scientific evidence (DNA, epidemiology, specific types of psychological evidence, including psychological syndromes and profiles) and their general techniques for managing scientific evidence in court (Gatowski et al., 2000). At the end of the telephone interview in Part I, the judge respondent was given the option of completing Part II on the telephone or having a questionnaire sent in the mail. Two equivalent versions of Part II were developed—one in a telephone survey format and one in a mail survey format. Part II was composed of mostly close-ended questions presented in an easy to follow format.

Securing Participation

Judges in Sample Replicate A were sent a letter of introduction outlining the nature, purpose, and goals of the research and the importance of the information to be obtained (Dillman, 1978, 1999; Dillman & Tarnai, 1988; Dobbin et al., 2001; Lavrakas, 1993). Approximately 10–14 days later, judges received a follow-up telephone call to solicit their participation in the study. Interview schedulers used a standardized script to further explain the importance of the research to the judiciary and to encourage participation. When making scheduling calls, staff were candid about the time commitment involved and always scheduled (and sometimes rescheduled) interviews at a time most convenient for the judge. Once a judge agreed to be interviewed, a time and date for the telephone interview was arranged. On average, it took four calls to make contact with most judges (range of 1–17 calls). The telephone survey in Part I took, on average, 55 min to complete. Judges did not receive any financial compensation for participation in the research.

Coding and Analysis of Survey Responses

An empirical code book was developed for both the close-ended and open-ended questions on the survey instruments. Codes for open-ended questions were developed based on a random drawing of 50 of the first 100 completed interviews. Responses to open-ended questions were reviewed and mutually exclusive codes were constructed for each question. The initial code book, developed on the 50 surveys, was then used to code responses on another 50 surveys, and revisions, additions, and deletions were made as necessary.

Coders were trained on appropriate coding and check-coding techniques. Coder reliability was assessed and maintained throughout production coding. Approximately 25% of all coding was check-coded (Cohen kappa = .84) and showed high, chance-corrected reliability. Frequencies and cross-tabulations were run on all variables of interest.

Assessment of the judge respondents' understanding of the specific *Daubert* guidelines received special attention. Their level of understanding was inferred from responses to a question asking judges how they would apply a particular *Daubert* criterion when evaluating the merits of proffered scientific evidence. Questions were open-ended and multiple responses were permitted. The entirety of the response was coded. Although a definition of the *Daubert* criterion was not provided as part of the question, if the judge asked for a definition the interviewer provided a scientifically correct, standardized, pretested definition and coded that the definition had been asked for and given. If a definition was not asked for by the respondent, it was not given by the interviewer.

The code book provided clearly articulated guidelines and instructions for coding level of understanding, and particular attention was paid to the coding of these responses (i.e., steps were taken to ensure high intercoder reliability; .84). Responses were assessed against predetermined criteria and then coded as "judge understands concept," "judge's understanding of concept is questionable," and "judge clearly does not understand concept." In order to be coded as "judge understands concept" for any *Daubert* criterion, the judge had to refer to the central scientific meaning of the concept. For example, with respect to falsifiability, in order for a response to be coded as "judge understands concept," the judge's response had to make explicit reference to testability, test and disproof, prove wrong a theory or hypothesis, or proof/disproof. If the judge did not explicitly refer to any of these central concepts, but instead appeared to talk around the issue, or just alluded to "a test of a theory or hypothesis," then the judge's understanding was coded as "questionable." A "questionable" meant that the coders were unable to confidently infer that the respondent truly understood the scientific meaning of the concept. It is recognized that this approach set a relatively high threshold for inferring respondent understanding of the *Daubert* guidelines.

RESULTS

Response Rates

A total of 400 surveys of judges were completed, with a response rate for Part I of the survey instrument of 71% (the telephone interview). Part II of the survey had an overall response rate of 81% (325 of 400 completed surveys in Part I). Of the 400 judges surveyed in Part I, 31% ($n = 123$) chose to complete Part II via telephone immediately after completion of Part I and 69% ($n = 276$) chose to complete Part II via mail. For judges who received Part II in the mail, there was a return rate of 73% (207 surveys returned of a possible 277). (See Dobbin, Gatowski, Ginsburg, Merlino, Dahir, & Richardson, 2001 for a more detailed discussion of steps taken to facilitate a high response rate).

Respondents' Characteristics

Admissibility Standards

Just over half of the judges surveyed were from states that followed *Daubert* and the Federal Rules of Evidence (FRE) ($n = 205$). The remainder of the sample ($n = 195$) were from states that followed *Frye* or some hybrid such as a *Kelly–Frye* standard used in California. Few significant differences were found between *Daubert* and non-*Daubert* respondents, so data for the two kinds of jurisdictions are collapsed except where indicated.

Time on the Bench

Half of the judges surveyed had been judges for more than 10 years ($n = 200$), with 31% of the judges ($n = 123$) reporting 5–10 years of experience on the bench, and 19% ($n = 77$) reporting less than 5 years on the bench.

Experience With Types of Evidence

Very few of the judges surveyed reported having any experience with epidemiological evidence, in fact the majority (73%, $n = 237$) reported no experience at all. Overall, almost two thirds of the judges (65%, $n = 210$) reported at least some experience with DNA evidence in their courtroom. More than three quarters of the judges surveyed (80%, $n = 260$) reported at least some experience with psychological evidence generally, with almost one quarter of the judges (22%, $n = 73$) reporting a great deal of experience with such evidence. Approximately three quarters of the judge respondents (74%, $n = 241$) reported at least some experience with testimony about the disorders contained within the *Diagnostic and Statistical Manual of Mental Disorders (DSM-IV)*, with one quarter of the judges (26%, $n = 83$) reporting a great deal of experience.

Science Education

Overall, judges were almost equally divided in their opinion of whether their education had adequately prepared them to deal with the range of scientific evidence proffered in their courtrooms. Fifty-two percent ($n = 209$) of the judges felt that they had been adequately prepared, whereas 48% ($n = 191$) believed their education had left them inadequately prepared. Since high school, but not counting continuing legal education (CLE) courses, most of the judges surveyed (85%, $n = 341$) reported that they had taken formal courses in the social sciences. Seventy-seven percent ($n = 307$) of the judges reported that they had taken formal courses in the physical sciences, and 67% ($n = 268$) had taken formal courses or training in the biological sciences. Although more than half of the judges (63%, $n = 251$) reported that they had received CLE training about the use of specific types of scientific evidence in the courtroom, the overwhelming majority of these judges (96%, $n = 241$ of 251) reported that they had *not* received instruction about general scientific methods and principles.

What do Judges Think About the Intent and Value of *Daubert*?

One third of the judges surveyed (32%, $n = 126$ of 400) believed that the intent of *Daubert* was to raise the threshold of admissibility for scientific evidence, whereas 23% ($n = 90$ of 400) believed that the intent was to lower the threshold for admissibility. Just over one-third (36%, $n = 142$ of 400) believed that the intent of the Supreme Court in *Daubert* was to neither raise nor lower the threshold for admissibility, but rather the intent was to articulate a framework for admissibility and to give judges the discretion to apply the guidelines as appropriate. The remaining judges (11%, $n = 43$ of 400) were uncertain as to the Supreme Court's intention. Most judges (75%, $n = 300$) believed that one purpose of *Daubert* was to guard against the admission of "junk science," with only 15% ($n = 6$ of 400) indicated that this was *not* an intended purpose. Ten percent ($n = 39$ of 400) were unsure. The belief on part of judges that one purpose of *Daubert* was to guard against "junk science" was significantly associated to their operating admissibility standard, with judges in *Daubert* states more likely to agree that this was an explicit intent of the decision ($p < .021$).

When asked what overall value the *Daubert* decision has for judicial decision-making with respect to the admissibility of proffered scientific evidence, just over half of the judges (55%, $n = 220$) reported that *Daubert* had a "great deal" of value. For these judges, the value of *Daubert* lay in its provision of a decision-making framework and the articulation of the "steps to consider when making admissibility decisions" and "the basis provided for justifying or explaining the decision-making process." Some of these judges considered the value of *Daubert* to be its potential for "greater consistency among the states with respect to admissibility decisions" and its ability to ensure that "junk science is kept out of the courtroom." Thirty-nine percent ($n = 156$) of the judges were less positive in their assessments of the value of *Daubert*, noting that it only had "some" value to their decision-making practice. Although these judges felt that *Daubert* provided a "good start at articulating a general framework for decision making," the guidelines themselves were "not precise or specific enough to be truly helpful." The few judges who reported that *Daubert* had "no value at all" (6%, $n = 24$) expressed concern that the guidelines' "lack of specificity had led to more confusion," that "*Daubert* had not really changed anything," and that "*Frye* is just as good." There was no significant difference in the perceived value of *Daubert* as a function of operating admissibility standard.

Do Judges See the Role of "Gatekeeper" as an Appropriate One?

All of the judges interviewed were asked how active a role they tend to take in determining the scientific merits of the proffered evidence. Perhaps not surprisingly, an overwhelming 91% of the 400 judges surveyed ($n = 364$) believed that the role of "gatekeeper" was an appropriate one for a judge, *irrespective* of the admissibility standard followed in their state. Almost two thirds of the judges surveyed (62%, $n = 249$) perceived themselves as taking an "active role" in the admissibility process (22% of respondents ($n = 87$ of 400) described themselves as taking a "very active role" and 41% ($n = 162$ of 400) described themselves taking a "somewhat active role"). In fact, almost all of the judges talked about the gatekeeping role as their fundamental "function," "what judges do," and a "necessary" role for the

judiciary. Those few judges who reported that the gatekeeping role was inappropriate (9%, $n = 36$), explained that judges' "lack of scientific training" made the performance of such a role "difficult," "untenable," and ultimately "inappropriate."

Judges from *Daubert* states ($n = 205$) were also asked if they thought their role with respect to admissibility decision-making had changed as a consequence of their state's adoption of *Daubert*; 52% ($n = 106$) believed that their judicial role had changed, 38% ($n = 78$) believed that their role had not changed, and 10% ($n = 20$) were unsure. Judges who felt that their role had changed believed that they had become more active gatekeepers under *Daubert*, regardless of whether they initially described their role as "very active," "somewhat active," or "minimally active." Common responses included, "[Under *Daubert*] the judge becomes more of a determiner of the reliability and validity of the evidence" and "The role of the judge is greater because of the added procedures that may occur under *Daubert*, such as pretrial admissibility hearings." Judges who reported *Daubert* had *not* changed their role in admissibility decision-making believed that under *Daubert* they were still doing what they had always been doing. As one judge commented, "*Daubert* put the judge back in the courtroom. Thing is, I've been there all along."

Can Judges Operationalize Daubert's Scientific Concepts and Appropriately Use Them as Decision-Making Guides?

Falsifiability

The majority of the judges surveyed (88%, $n = 352$), regardless of the admissibility standard followed in their state, believed falsifiability to be a useful guideline for determining the merits of proffered scientific evidence. In fact, only 8% ($n = 32$) of the judges surveyed reported that falsifiability was "not at all useful" as a decision-making tool.

As a follow-up to the question about the utility of the guideline, judges were asked to discuss how they would apply the guideline of "falsifiability" in determining the admissibility of proffered scientific evidence (i.e., "How would you use the guideline of falsifiability when scientific expert testimony is proffered in your court?"). The question was designed in such a way as to allow the researchers to infer how well the judge actually understood the scientific meaning of the criterion. Multiple responses to open-ended questions were coded and some examples of the most common responses for each of the coding categories are provided for illustrative purposes. Although the majority of judge-respondents reported falsifiability to be useful when determining the merits of proffered scientific evidence, the results clearly indicate that most judges did not fully understand the scientific meaning of this concept.

From the answers that were provided, the researchers could only infer a true understanding of the scientific meaning of falsifiability in 6% ($n = 23$ of 400) of the judges' responses. In fact, for the 352 judges who indicated that falsifiability was a useful criterion, the coders could only infer a true understanding of the concept in 4% of the responses elicited. Responses such as, "I would want to know to what extent the theory has been properly and sufficiently tested and whether or not there has been research that has attempted to prove the theory to be wrong" and "if it is not possible to test the evidence then it would weigh heavily with me in my decision"

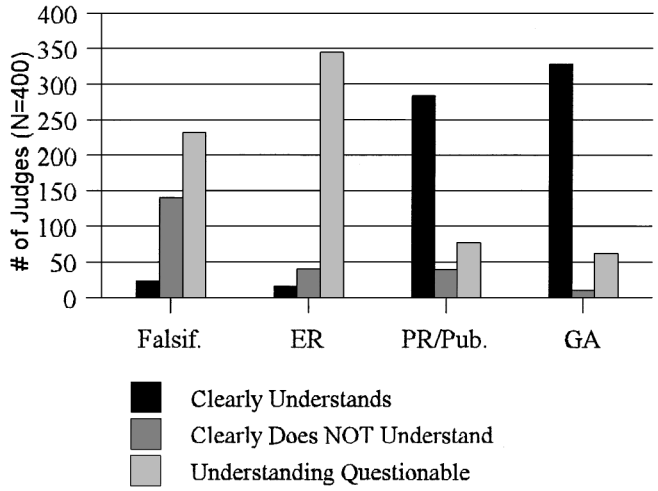

Fig. 1. Understanding × Guideline.

are illustrative of responses coded as "judge understands concept." An example of a response coded as "questionable" would be, "I would want to know if the theory has been tested," without further articulation of what this means. Just over one third of the judges (35%, $n = 140$ of 400) provided a response that clearly indicated that the judge did *not* understand the scientific meaning of falsifiability (e.g., "I would want to know if the evidence was falsified," "I would look at the results and determine if they are false"). Interestingly, despite the general sense of hesitancy in providing a response regarding how they would use the criterion of falsifiability when making an admissibility decision, only 16% of the judges surveyed asked for a definition or further explication of the guideline before attempting an answer. (See Table 1 and Figure 1 for a summary of these results).

Error Rate

The vast majority of judges (91%, $n = 364$) indicated that a consideration of error rate was useful when determining the merits of proffered scientific evidence, with just over half (54%, $n = 214$) stating that it was "very useful."

Despite the general agreement that examining error rates when making an admissibility decision was useful, once again the responses revealed a general lack of understanding of the scientific meaning of error rate. In order for a response to be coded as "judge understands concept" the response had to include reference to false positives and false negatives, or refer to a number or percent of instances in which a classification procedure led to misclassification. Mere reliance on a high/low heuristic without further articulation was not sufficient. The following responses are illustrative of those coded as "judge understands concept": "it would seem that if a theory or procedure has too high an error rate it would have to be rejected because the risk is too high of being wrong . . . ," "I would want to know about the probability of making a mistake." When asked a question about how they would apply the concept of error

Table 1. Utility of *Daubert* Guidelines × Operating Admissibility Standard[a]

		Falsifiability	ER	PR & Pub.	Gen. Acct.
"Very useful"	*Daubert* ($n = 205$)	35% ($n = 71$)	51% ($n = 105$)	56% ($n = 114$)	59% ($n = 120$)
	Frye ($n = 195$)	42% ($n = 82$)	56% ($n = 110$)	48% ($n = 94$)	69% ($n = 134$)
"Somewhat useful"	*Daubert* ($n = 205$)	57% ($n = 116$)	42% ($n = 86$)	37% ($n = 76$)	36% ($n = 74$)
	Frye ($n = 195$)	43% ($n = 83$)	32% ($n = 63$)	43% ($n = 84$)	22% ($n = 43$)
"Not at all useful"	*Daubert* ($n = 205$)	6% ($n = 13$)	5% ($n = 11$)	5% ($n = 11$)	3% ($n = 7$)
	Frye ($n = 195$)	10% ($n = 19$)	6% ($n = 11$)	7% ($n = 13$)	4% ($n = 7$)

Note: ER = Error rate; PR & Pub. = Peer review and publication; Gen. Acct. = General acceptance.
[a]There was no significant difference in perceived utility of *Daubert* guidelines as a function of the operating admissibility standard.

rate to a determination of admissibility, a clear understanding was revealed in only 4% ($n = 16$) of the responses. For the 364 judges who indicated that error rate was a useful criterion, the coders could only infer a true understanding of the concept in 4% (15 of 364) of the responses provided. In 86% of the responses ($n = 344$ of 400) the judges' understanding of the concept was questionable at best (e.g., "If the error rate is high, I would take that into account"), and in 10% of responses ($n = 40$ of 400) judges clearly had little understanding of the scientific meaning of error rate (e.g., "I would take into account the number of mistakes that were made and consider that in my admissibility decision"). And, despite the general sense of hesitancy in providing a response regarding how they would use the criterion of error rate when making an admissibility decision, only one judge asked for a definition or further explication of the guideline before providing an answer. (See Table 1 and Figure 1 for a summary of these results).

Peer Review and Publication

The majority of the judges surveyed (92%, $n = 368$), regardless of admissibility standard followed in their state, felt that the concept of peer review was useful for determining the admissibility of expert evidence, with just over half (52%, $n = 208$) reporting that it was a "very useful" guideline. Only 6% ($n = 24$) of judge-respondents overall indicated that peer review was "not at all useful" in determining the admissibility of scientific evidence.

The majority of judges noted that they would be highly likely to reject anything not subjected to rigorous peer review analysis, and comments such as "substantial weight should be given to peer review as it gives the evidence credibility" were frequent. When asked how they might apply the concept of peer review to a determination of the admissibility of proffered evidence, most judges (71%, $n = 284$ of 400) provided responses that demonstrated a clear understanding of the scientific peer review process. Some examples of actual responses elicited: "Peer review gives you an idea of whether this is a scientific idea that has been debated in the field. It would be important to have the experts describe and debate the peer-reviewed literature to determine the acceptability of the evidence"; and "I would give greater weight to a criticism of the technique or procedure if it appeared in a significant number of high status journals, or if the prestige of the criticizer was high ... nevertheless, practically speaking, it would be difficult for me to evaluate the prestige of the critic." Only 10% of the respondents ($n = 39$) gave a response that clearly reflected a lack of understanding with respect to the application of peer review and publication in the decision-making process. Five percent of the respondents asked for a definition of "peer review and publication" before providing a response. (See Table 1 and Figure 1 for a summary of these results).

General Acceptance

Not surprisingly, the vast majority of judges (93%, $n = 371$), regardless of operating admissibility standard, indicated that general acceptance was a useful criterion for determining the merits of the proffered scientific evidence, with 64% ($n = 254$) indicating that it was a "very useful" guideline. Again, not surprisingly given its relation to the *Frye* standard, the majority of judges (82%, $n = 328$) also demonstrated

a clear understanding of the concept of general acceptance when asked to discuss how they would apply the guideline to the admissibility of expert evidence. In fact, no judge asked for a definition of general acceptance before providing a response.

When asked how they would determine the relevant scientific fields within which to determine the general acceptance of a theory or technique, the majority of respondents (74%, $n = 295$ of 400) indicated that they would hear expert testimony on the issue, one quarter of the judges (24%, $n = 97$ of 400) indicated that they would make that determination on a "case-by-case basis," and one-fifth (19%, $n = 76$ of 400) indicated that they would make that determination based upon the relevance of the evidence to the issue at hand. Fifty-six of the judges (14%) indicated that they would rely on their own knowledge to make such a determination. (See Table 1 and Figure 1 for a summary of these results).

Do Judges Attribute Equal Importance to Each of the Guidelines?

Judges were asked to what extent, if any, and under what circumstances they would weight or combine the four guidelines of falsifiability, error rate, peer review and publication, and general acceptance when determining the admissibility of proffered expert evidence. Overall, the survey found little consensus regarding how to weight or combine the *Daubert* criteria.

Twenty percent of the judges surveyed ($n = 80$ of 400) reported that they would weight or combine the *Daubert* criteria on a case-by-case basis depending upon the particular type of evidence being proffered and which criterion was most helpful in that instance. One hundred and sixty-six of the judges (42%) chose to answer the question by indicating the *Daubert* guideline to which they would generally attribute the *most* weight—of these judges, half (86 of 166, or 22% of 400) indicated that general acceptance would be given the most weight, 18% (30 of 166, or 8% of 400) indicated that falsifiability would be given the most weight, 16% (26 of 166, or 7% of 400) indicated that error rate would be given the most weight, and 14% (24 of 166, or 6% of 400) indicated that peer review and publication would be given the most weight. Seventeen percent of the judges surveyed ($n = 69$ of 400) indicated that *all* the criteria have equal weight, and 21% ($n = 85$ of 400) were unsure about how to combine the four guidelines.

Do Judges Apply the Guidelines Differently to Different Domains of Knowledge and Does the Judicial "Bench Philosophy of Science" Reflect the Philosophical Assumptions of Daubert?

Whether judges believed that they could distinguish "scientific knowledge" from "technical or otherwise specialized knowledge" was not significantly associated with the operating standard of admissibility in their states. Almost two thirds of the judges surveyed (61%, $n = 243$ of 400) believed that "scientific knowledge" could be distinguished from "technical or otherwise specialized knowledge." These respondents ($n = 243$) were then asked to discuss how they would make that distinction.

The majority of these judges (84%, $n = 203$ of 243) indicated that the distinction between "science" and some other form of knowledge should be made on a

case-by-case basis, depending on the nature of the evidence proffered, the purpose for which the evidence is proffered, the qualifications of the expert offering the evidence, and existing precedents. Most of the judges who believed that a distinction could be made between "science" and other forms of knowledge (60%, $n = 145$ of 243) contrasted scientific knowledge as the "generation of new knowledge" with technical knowledge as the "application of known facts." Thirty-one percent of judges ($n = 76$ of 243) mentioned that "science is objective and less open to interpretation."

One third of all of the judges surveyed ($n = 130$ of 400) said that "scientific" knowledge could *not* be distinguished from other forms of knowledge for the purpose of deciding admissibility. Moreover, approximately half (54%, $n = 70$ of 130) indicated that not only is the distinction not helpful, but that it is also "not relevant to the admissibility decision." Twenty percent of these judges ($n = 26$ of 130) reported that the more important question is one of "fit" or the "relevance of the expert evidence to the facts at issue." Twenty percent ($n = 26$ of 130) also believed that deciding whether expert evidence was "scientific" was a question better suited for the jury and not for the judge. Nineteen percent ($n = 25$ of 130) indicated that from the perspective of the judge on the bench and the admissibility decision, there is "no real substantive difference between scientific knowledge and technical or otherwise specialized knowledge." Seven percent of all the judges surveyed ($n = 27$ of 400) were "unsure" whether a distinction could be made between "scientific knowledge" and "technical and otherwise specialized knowledge."

All of the judges who believed they could differentiate between scientific knowledge and other knowledge ($n = 243$ of 400) were asked to designate specific domains of expert testimony as either "scientific evidence" or as "technical or otherwise specialized knowledge," and to articulate their reasons for making such a designation. Specific areas of social science evidence were selected for focus (psychological testimony based on clinical assessments, psychological testimony based on laboratory studies, survey evidence, economic analyses, and engineering analyses). These results are summarized in Table 2.

Table 2. Characterization of Specific Forms of Evidence as "Scientific" "Technical/Specialized" Knowledge[a]

	Scientific	Technical/specialized	DK
Psychological testimony derived from clinical inference	38% ($n = 93$)	60% ($n = 147$)	1% ($n = 3$)
Psychological testimony derived from laboratory study	64% ($n = 156$)	31% ($n = 75$)	5% ($n = 12$)
Survey evidence, such as might be presented in a case about trademark infringement	12% ($n = 28$)	85% ($n = 206$)	4% ($n = 9$)
Economic evidence, such as might be presented to argue for loss of earning in a personal injury case	17% ($n = 42$)	83% ($n = 201$)	— —
Engineering evidence, such as an analysis of a structural design defect	56% ($n = 135$)	33% ($n = 81$)	11% ($n = 27$)

[a]Total $n = 243$ (61 of the sample of 400). All reported that a distinction could be made between scientific and technical and other specialized knowledge.

Expert Evidence Based on Psychological Testimony Derived From a Clinical Inference

Overall, more than half of the judges who felt a distinction between scientific and other domains of knowledge was possible classified evidence based on a clinical psychological inference (e.g., psychological assessment and diagnosis) as "technical or otherwise specialized knowledge" (60%, $n = 147$ of 243), whereas approximately one-third (38%, $n = 93$ of 243) defined such testimony as "science."

The majority of judges, regardless of operating admissibility standard, who designated testimony based on psychological clinical inference as "science," (86%, $n = 80$ of 93) relied upon "general acceptance" as the justification for their designation. Ten percent of judges in *Daubert* states designating the evidence as "science" ($n = 5$ of 48) commented that it could be "falsified," whereas 17% of judges from *Daubert* states designating the evidence as "technical or otherwise specialized knowledge" justified their designations on the basis that this type of psychological evidence "could not be falsified" ($n = 13$). For those judges designating the evidence as "technical or otherwise specialized knowledge, the majority (84%, $n = 124$ of 147), regardless of admissibility standard, referred to the "overly subjective nature" of such evidence. Only four judges explicitly mentioned peer review and publication as a justification for their designation, and none of the judges mentioned error rates.

Expert Evidence Based on Psychological Testimony Derived From Laboratory Studies

With respect to psychological testimony derived from laboratory studies (e.g., eyewitness research), most of the judges (64%, $n = 156$ of 243) designated such evidence as "science," whereas 31% ($n = 75$ of 243) designed such evidence as "technical or otherwise specialized knowledge."

Almost two-thirds of the judges surveyed (64%, $n = 156$ of 243) designated psychological laboratory studies as "science." Of these judges, 67% justified their designation on the grounds that such evidence has reached a standard of "general acceptance." One third of these judges ($n = 52$ of 156) noted that psychological laboratory studies are "derived from the scientific method." Only 17% ($n = 26$ of 156) of the judges referred to falsifiability in justifying their "science" designation. There was also very little attention paid to error rate analysis and the peer review and publication process in designating testimony based on psychological laboratory studies as "scientific" or "other" knowledge. The majority of judges who designated such evidence as "technical or otherwise specialized knowledge" (84%, $n = 63$ of 75), referred to the "subjective nature of the evidence" as grounds for their designation. Interestingly, 23% of the judges indicated that such evidence did not match what they generally thought of as being "science."

Expert Evidence Based on a Survey

The majority of judges (85%, $n = 206$ of 243) designated evidence based on survey research (e.g., such as research presented in a case about trademark infringement) as "technical or otherwise specialized knowledge," whereas 12% ($n = 28$ of

243) designated such evidence as "scientific knowledge." General acceptance was the most commonly noted justification for designating expert evidence based on a survey as "scientific" (82%, $n = 23$ of 28).

Expert Evidence Based on Economic Analyses

The majority of judges drawing a distinction between "scientific" and "technical or otherwise specialized knowledge" (83%, $n = 201$ of 243) designated evidence based on an economic analysis as "technical or otherwise specialized knowledge," whereas 17% ($n = 42$ of 243) designated such evidence as "science."

Expert Evidence Based on Engineering

Just over half of the judges who distinguished "scientific" knowledge from other forms of knowledge (56%, $n = 135$ of 243) designated expert evidence derived from engineering as "science," whereas 33% ($n = 81$ of 243) designated such evidence to be "technical or otherwise specialized knowledge."

The majority of judges designating engineering evidence as "science" relied upon general acceptance for their justification (81%, $n = 105$ of 135). In contrast, 41% of the judges who designated the evidence as "technical or otherwise specialized knowledge" ($n = 33$ of 81) noted a "lack of general acceptance" as a reason for their designation. Nineteen percent of the judges designating the evidence as "science," commented that such evidence "could be falisified" and 10% noted that the "inability to falsify" such evidence justified a designation of "technical and otherwise specialized knowledge."

DISCUSSION

The vast majority of judges surveyed believed that the role of gatekeeper was an appropriate one, irrespective of the admissibility standard followed in their state. Moreover, not only was the role seen as an appropriate one, but the judges reported that they actively engaged in their gatekeeping function. It is interesting to note that the few judges who felt the gatekeeping role was inappropriate were not commenting on the gatekeeping role per se, but rather on their perceived lack of ability to perform that role due to a lack of sufficient background and training in scientific methods and principles. Recall that the judges were almost equally divided in their opinion about whether their education, including continuing legal and judicial education, had adequately prepared them to deal with the range of scientific evidence proffered in their courtrooms. Although judges were divided on whether the intent of *Daubert* was to raise or lower the threshold for admissibility, the belief that one purpose of *Daubert* was to guard against "junk science" was significantly associated with the admissibility standard operating in their state, with judges in *Daubert* states more likely to agree that this was an explicit intent of the decision.

As results of our national survey clearly indicate, judges believe that the *Daubert* criteria are useful guidelines for determining the admissibility of proffered expert evidence. Not only did the judges surveyed find the individual criteria useful, but

many of the judges reported that the value of the *Daubert* standard itself lay in its provision of a decision-making framework and articulation of guiding criteria. However, consistent with our expectations, although the judges surveyed reported that they found the *Daubert* criteria useful for determining the admissibility of proffered expert evidence, the extent to which judges understand and can properly apply the criteria when assessing the validity and reliability of proffered scientific evidence was questionable at best.

The survey findings strongly suggest that judges have difficulty operationalizing the *Daubert* criteria and applying them, especially with respect to falsifiability and error rate. Only a very small percentage of judges surveyed provided responses that clearly reflected an understanding of the *scientific* meaning of those two criteria. Most of the respondents talked around the concepts and offered only passing, if any, reference to their central meaning. This finding is also supported by case law reviews that have analyzed judicial opinions for discussions of the *Daubert* criteria (Groscup et al., 2000; Richardson, Dobbin, Gatowski, Ginsburg, Merlino, Dahir, & Colton, 1998; Richardson, Ginsburg, et al., 1995).

It is interesting to note the high proportion of responses for which degree of understanding for falsifiability and error rate was unclear (i.e., those that were coded as "judge's understanding questionable"). Judges who gave an ambiguous response were probed by the interviewer to elicit further discussion with the aim of getting the judge to talk through his or her unarticulated understanding. Yet, despite these efforts, the majority of judges were unable to provide an answer that clearly demonstrated their understanding of the concepts. It may be that the judges did understand the concept and their ambiguity is an artifact of the way in which the question was asked (recall that understanding was being inferred from a question that asked how they would apply the concept when making an admissibility decision), or that they understand the concept but were unable to articulate their understanding. On the other hand, it seems likely that the ambiguity of the responses may reflect a genuine lack of understanding of these scientific concepts.

Although we recognize that the standard in this research against which judges' level of understanding was measured was high, and that the general nature of the question may have presented somewhat of an artificial context for the judge-respondent (i.e., the respondents were not asked to make a judgment based on the presentation of specific facts in a specific case), one can argue that the gatekeeping role prescribed by *Daubert* assumes, if not necessitates, a general level of judicial scientific literacy. Although hesitancy on the part of some judge-respondents in discussing how they might apply a particular criterion when making an admissibility decision may reflect some discomfort with their level of understanding, or at least their ability to articulate their understanding of the concept, only a small percentage asked the interviewer for a definition of the term or for further explanation. Thus, although the judges confidently and overwhelming responded that the *Daubert* criteria were useful decision-making guides, the majority did not seem to recognize or acknowledge their lack of understanding about how to apply some of the guidelines as part of the admissibility decision-making process.

This is an important finding. Judges' difficulty operationalizing the *Daubert* criteria, especially falsifiability and error rate, suggests limitations in the judiciary's

understanding of science. This finding questions the ability of the courts, particularly the state trial courts, to assess the scientific reliability and validity of proffered scientific evidence and hints at the potential for inconsistencies in *Daubert*'s application, especially after the *Kumho* decision (Saks, 2000). Judges' lack of sophistication regarding the scientific meaning of the *Daubert* criteria should cause concern about whether judges are making accurate and reliable assessments of proffered scientific evidence using the *Daubert* criteria. Although the value of *Daubert* may lie, according to the judges surveyed, in its provision of a decision-making framework, a general lack of understanding of the scientific meaning of the guidelines and an inability to operationalize them would seem to undermine this value. Thus, the practical value of *Daubert* for judges may never be fully realized unless judges are provided with sufficient judicial scientific education to allow them to perform their gatekeeping role.

When judges were asked to what extent, if any, and under what circumstances they would weight and combine the four guidelines, there was little consensus. Although most judges agreed that a distinction could be made between "scientific" and "technical or otherwise specialized" knowledge, their ability to apply the *Daubert* guidelines seemed to have little bearing on whether specific types of expert evidence were designated as "science" or "nonscience." An examination of the justifications given for designating certain types of contested evidence as "scientific" or "technical and otherwise specialized knowledge" reveals that general acceptance was the most frequently relied upon guideline. It is important to recognize that this finding may be attributable, in part, to the nature of the evidence included in the study (i.e., psychological clinical inference, psychological laboratory studies, survey data, economic analyses, and engineering). This type of evidence often sits on the contested boundary of what is and is not considered "science" within the legal arena. However, it is also important to note that even for those judges who designated technical or otherwise specialized types of evidence as "science," there was very little mention of any *Daubert* guideline, other than general acceptance, as justification for their designations.

For those judges who did mention the criteria as justification for a designation, their responses reflected more of the rhetoric of *Daubert* than the substance. In some cases, there was a reference to a guideline (e.g., falsifiability) to support a designation of "science" and a reference to the same guideline by other judges to support a designation of the same evidence as "technical or otherwise specialized knowledge." Again, not only does this suggest that most judges have difficulty operationalizing some of the *Daubert* concepts, but it also suggests that application of the concepts such as falsifiability and error rate may prove particularly difficult to certain types of contested, "less scientific" domains of knowledge. It also serves to highlight the potential for inconsistent application of the *Daubert* guidelines. However, the lack of consensus on how and when to apply the guidelines, alone or in combination, might also reflect the judges' understanding of their gatekeeping role, the discretion inherent in that role (especially after *Joiner* and *Kumho*), and the explicitly "flexible" nature of the *Daubert* guidelines.

Implications for the Evolving Relationship Between Science and Law. The results of the national survey indicate that although judges overwhelmingly endorse

the active gatekeeping role defined by *Daubert*, many may lack the scientific literacy necessitated by *Daubert*. Although the judges clearly found the *Daubert* criteria to be useful decision-making guidelines, many had limited knowledge of how to use some of the guidelines in practice. Moreover, there was no significant difference in the perceived value of *Daubert*, in the degree of understanding of the *Daubert* guidelines, or in the degree to which they are relied upon in the decision-making process, between judges in *Daubert* and non-*Daubert* states.

These findings raise issues of policy, practice, and philosophy. The judges who commented that for the purposes of admissibility there is no substantive difference between "scientific knowledge" and "technical or otherwise specialized knowledge" foreshadowed the Supreme Court's direction in *Kumho*. In *Kumho*, the Supreme Court expanded the scope of *Daubert* to include all forms of expert knowledge and moved the court away from the necessity of distinguishing between "scientific" and "technical or otherwise specialized knowledge." However, although on its surface *Kumho* resolved one of the central debates of *Daubert* and its application to different forms of knowledge, the decision failed to address the underlying assumption that judges are fully capable of making judgments about the scientific reliability and validity of proffered scientific evidence. In fact, because bench philosophies of science—judicial definitions of what constitutes science—seem to reflect the rhetoric but not the substance of *Daubert*, *Kumho* may ultimately have clouded the process even further. *Kumho* explicitly strengthened the discretion given to judges by *Daubert* in noting that judges had "broad latitude" in deciding to apply the guidelines to different forms of expert knowledge (*Kumho*, at 1176). The survey findings raise concerns, however, about how well judges can exercise their discretion if they lack the requisite understanding of science and its methods. That is, if judges do not fully understand what falsifiability means, for example, then how are they to decide whether to apply it to a specific form of expert evidence and then how to apply it appropriately? Moreover, if consideration of the *Daubert* criteria is problematic for "traditional" forms of scientific evidence, the consideration of these criteria becomes even more difficult when evaluating other forms of expert knowledge, as well as scientific knowledge that may not easily fit in the *Daubert* framework (e.g., some forms of social and behavioral science). This is not to suggest that many forms of contested science should not be held to a high standard of admissibility. But any standard of admissibility, with its inherent flexibility, has to be fully understood so that it can be properly applied. To give judges discretion without the proper foundation from which to exercise that discretion creates the potential for inconsistent decisions (Saks, 2000) and for a mechanistic application of the *Daubert* criteria, which in itself reflects a misunderstanding of the foundations of science and its methods (e.g., probabilistic reasoning and statistical analysis, the cumulative nature of scientific knowledge, and the sociopolitical contexts within which scientific knowledge is generated and disseminated) (see also Gatowski, Dobbin, Richardson, & Ginsburg, 1997; Jasanoff, 1993; Richardson, Gatowski, & Dobbin, 1995).

Given the active gatekeeping role that judges report taking in admissibility decision-making, even in states that do not explicitly follow *Daubert*, the general lack of scientific literacy among the state trial judiciary, and the increasingly complex nature of the science that comes before the court, the research presented herein

clearly demonstrates the need for more science-based judicial education. In recent years a number of educational resources have been developed to assist judges in understanding their gatekeeping role and to help them properly apply the appropriate admissibility standards in the courtroom (cf. Dobbin & Gatowski, 1999; Faigman et al., 1997, 1999; Federal Judicial Center, 1994, 2000; Parry, 1998). The question becomes, then, what is the appropriate scope and purpose of science-based judicial education?

What judges need to know is not how to design the best scientific study, but how to evaluate imperfect ones. Judges do not need to be trained to become scientists, they need to be trained to be critical consumers of the science that comes before them. This is an important distinction. It is inappropriate to suggest, we would argue, that judges should have the level of scientific literacy necessary to either design a scientific study or analyze complex statistics, nor is it appropriate to assume that judges should be able to review a scientific report or article and critique its methodology and conclusions with the same degree of rigor as a well-trained scientist. These tasks are not consistent with the judicial role of gatekeeper, nor do they recognize the complexity of the decision-making process. What judges do need is to become critical consumers of proffered expert evidence. Judges need to know what critical questions to ask, they need to know what methodological and statistical issues scientific experts, and other purveyors of science, should address and comment on when proffering science for use in the court. Judges need to know what to listen and look for when expert evidence is presented and what they should be asking about when the information is not forthcoming. Determining just what constitutes a sufficient level of scientific understanding for the judiciary is a question for future study and policy development.

The evolving relationship between science and law is bigger than *Frye*, *Daubert*, *Joiner*, and *Kumho*. Indeed, the lack of significant difference in the survey findings as a function of different operating admissibility standards is, in itself, a "significant" finding. The ongoing debates around the interface between science and law are pervasive and cross-state statutory boundaries and federal districts. This evolving relationship also has implications for the disciplines and practice of law and science. Indeed, expert evidence for use in the court is becoming, one might argue, an industry unto itself (Faigman, Kaye, Saks, & Sanders, 2000; Richardson, Gatowski, et al., 1995). Those involved in legal education at every level should make efforts to raise the scientific literacy of all those involved in the legal system. And, as the court sets boundaries on what will and will not be considered admissible expert evidence, scientific or otherwise, the scientific community will have to respond and hold itself accountable for the rigor of its research and the proffer of its evidence.

ACKNOWLEDGMENTS

This research was supported by a grant from the State Justice Institute (SJI-97-02B-D-030). Additional research support was provided by The Federal Judicial Center, The National Judicial College, The Grant Sawyer Center for Justice Studies at the University of Nevada, Reno, The Judicial Studies Program at the University

of Nevada, Reno, the National Council of Juvenile and Family Court Judges, and the Office of Juvenile Justice and Delinquency Prevention, U.S. Department of Justice (Grant No. 96-CT-NX-001). Considerable thanks are owed to our research staff whose willingness to conduct interviews at all times of day and commitment to a high standard of practice in data collection, coding, and analysis was greatly appreciated. We also thank our national advisory committee for their helpful advice and for the time spent reviewing instrumentation. And last, but certainly not least, we especially thank all of the judges who took time out of their busy schedules to participate in this research effort and provide us with their thoughts on this important topic.

REFERENCES

Agrimonti, L. M. (1995). The limitations of *Daubert* and its misapplication to quasi-scientific experts. *Washburn Law Journal*, *35,* 134–156.

Beecher-Monas, E. (1999). A ray of light for judges blinded by science: Triers of science and intellectual due process. *Georgia Law Review*, *33,* 1047–1111.

Berger, M. (1994). Procedural paradigms for applying the *Daubert* test. *Minnesota Law Review*, *78,* 1345–1386.

Daubert v. Merrell Dow Pharmaceuticals, Inc., 509 U.S. 579 (1993).

de Vyver, K. I. (1999). Comment: Opening the door but keeping the lights off: Kumho Tire Co. v. Carmichael and the applicability of the *Daubert* test to non-scientific evidence. *Case Western Reserve Law Review*, *50,* 177–202.

Dillman, D. A. (1978). *Mail and telephone surveys: The total design method.* New York: Wiley.

Dillman, D. A. (1999). *Mail and internet surveys: The tailored design method.* New York: Wiley.

Dillman, D. A., & Tarnai, J. (1988). Administrative issues in mixed mode surveys. In R. M. Groves et al. (Eds.), *Telephone survey methodology* (pp. 509–528). New York: Wiley.

Dobbin, S. A., & Gatowski, S. I. (1999). *A judge's deskbook on the basic philosophies and methods of science* [On-line]. The State Justice Institute. Available: www.unr.edu/bench

Dobbin, S. A., Gatowski, S. I., Ginsburg, G. P., Merlino, M. L., Dahir, V., & Richardson, J. T. (2001). Surveying difficult populations: Lessons learned from a national survey of state trial court judges. *Forthcoming, Justice System Journal*, *22*(3).

Dwyer, D. (1994). Federal rules of evidence supercede general acceptance standard for admissibility of scientific evidence. *Suffolk University Law Review*, *28,* 252–260.

Faigman, D. L. (1995). The evidentiary status of social science under *Daubert*: Is it "scientific," "technical," or "other" knowledge? *Psychology, Public Policy and Law*, *1,* 960–979.

Faigman, D. L., Kaye, D. H., Saks, M. J., & Sanders, J. (Eds.). (1997). *Modern scientific evidence: The law and science of expert testimony* (Vols. 1–2). St. Paul, MN: West Publishing Company.

Faigman, D. L., Kaye, D. H., Saks, M. J., & Sanders, J. (Eds.). (1999). *Modern scientific evidence: The law and science of expert testimony* (Vol. 3). St. Paul, MN: West Publishing Company.

Faigman, D. L., Kaye, D. H., Saks, M. J., & Sanders, J. (2000). Reply essay: How good is good enough?: Expert evidence under *Daubert* and *Kumho*. *Case Western Reserve University Law Review*, *50,* 645–647.

Faigman, D. L., & Wright, A. J. (1997). The battered woman syndrome in the age of science. *Arizona Law Review*, *39,* 67–115.

Federal Judicial Center. (1994). *Reference manual on scientific evidence*. Washington, DC: Federal Judicial Center.

Federal Judicial Center. (2000). *Reference manual on scientific evidence* (2nd ed.). Washington, DC: Federal Judicial Center.

Fenner, G. M. (1996). The *Daubert* handbook: The case, its essential dilemma, and its progeny. *Creighton Law Review*, *29,* 939–1089.

Frye v. United States, 293 F. 1013 (D.C. Cir. 1923).

Gatowski, S. I., Dobbin, S. A., Johnson, M. T., Cecil, J. S., Dahir, V., Merlino, M. L., Richardson, J. T., & Ginsburg, G. P. (2000). *The experience of state trial court judges with respect to expert testimony: Problems encountered and solutions employed*. Paper presented at the meeting of the American Psychology-Law Association (Div. 41, American Psychological Association), New Orleans, LA.

Gatowski, S. I., Dobbin, S. A., Richardson, J. T., & Ginsburg, G. P. (1997). The globalization of behavioral science evidence about battered women: A theory of production and diffusion. *Behavioral Sciences and the Law*, *15,* 285–305.
General Electric Co. v. Joiner, 522 U.S. 136 (1997).
Goodman-Delahunty, R. (1997). Forensic psychological expertise in the wake of *Daubert*. *Law and Human Behavior*, *21,* 121–140.
Gottesman, M. H. (1994). Admissibility of expert testimony after *Daubert*: The prestige factor. *Emory Law Journal*, *43,* 867–884.
Groscup, J. L., Studebaker, C. A., Huss, M. T., O'Neil, K. M., & Penrod, S. D. (2000). *A 'scientific' analysis of the admissibility of expert testimony: Investigating the effects of Daubert v. Merrell Dow Pharmaceuticals*. Paper presented at the meeting of the American Psychology-Law Association (Div. 41, American Psychological Association), New Orleans, LA.
Grove, W. M., & Barden, R. C. (1999). Protecting the integrity of the legal system: The admissibility of testimony from mental health experts under *Daubert*/*Kumho* analyses. *Psychology, Public Policy, and Law*, *5*(1), 224–242.
Hawthorne Partners v. AT&T Technologies Inc., No. 91 C 7167, 1993 WL 311916 (N.D. Ill. Aug. 11, 1993).
Iacobelli Construction, Inc., v. County of Monroe, 32 F. 3d 19 (2d Cir. 1994).
In re Aluminum Phosphide Antitrust Litigation, 893 F. Supp. 1497 (D. Kan. 1995).
Isely v. Capuchin Province, 877 F. Supp. 1055 (E.D. Mich. 1995).
Jasanoff, S. (1993). What judges should know about the sociology of science. *Judicature*, *77*(2), 77–82.
Krauss, D. A., & Sales, B. D. (1999). The problem of "helpfulness" in applying *Daubert* to expert testimony: Child custody determinations in family law as an exemplar. *Psychology, Public Policy, and Law*, *5*(1), 78–99.
Kumho Tire Co. v. Patrick Carmichael, 119 S. Ct. 1167 (1999).
Laser, J. (1997). Note: Inconsistent gatekeeping in federal courts: Application of Daubert v. Merrell Dow Pharmaceuticals, Inc. to nonscientific expert testimony. *Loyola of Los Angeles Law Review*, *30,* 1379–1420.
Lavrakas, P. J. (1993). *Telephone survey methods: Sampling, selection, and supervision*. Newbury Park, CA: Sage.
Moore v. Ashland Chemical, Inc., 126 F.3d 679 (5th Cir. 1997).
Needham, K. L. (1998). Note: Questioning the admissibility of nonscientific evidence after *Daubert*: The need for increased judicial gatekeeping to ensure the reliability of all expert testimony. *Fordham Urban Law Journal*, *25,* 541–574.
Odgers, S., & Richardson, J. T. (1995). Keeping bad science out of the courtroom: Changes in Australian and American expert evidence law. *University of New South Wales Law Review*, *18,* 108–129.
Parry, J. W. (1998). *National benchbook on psychiatric and psychological evidence and testimony*. ABA Commission on Mental and Physical Disability Law and The State Justice Institute.
Penrod, S., Fulero, S., & Cutler, B. (1995). Expert psychological testimony on eyewitness reliability before and after *Daubert*: The state of law and science. *Behavioral Sciences and the Law*, *13,* 229–259.
Petruzzi's IGA Supermarkets, Inc., v. Darling-Delaware Co., 998 F.2d 1224 (3d Cir 1993).
Renaker, T. S. (1996). Evidentiary legerdemain: Deciding when *Daubert* should apply to social science evidence. *California Law Review*, *84,* 1657–1692.
Richardson, J. T. (1994). Dramatic changes in American expert evidence law: From *Frye* to *Daubert*, with special attention to implications for social and behavioural science evidence. *Judicial Review*, *2,* 13–37.
Richardson, J. T., Dobbin, S. A., Gatowski, S. I., Ginsburg, G. P., Merlino, M. L., & Dahir, V. (1998). *Judges' experience with, and perception of, Daubert, scientific evidence, and the question of admissibility: A national study*. Paper presented at the meeting of the Law & Society Association, Aspen, CO.
Richardson, J. T., Dobbin, S. A., Gatowski, S. I., Ginsburg, G. P., Merlino, M. L., Dahir, V., & Colton, C. (1998). *A case law survey of social and behavioral science evidence after Daubert*. Paper presented at the meeting of the American Psychology-Law Association (Div. 41, American Psychological Association), Redondo Beach, CA.
Richardson, J. T., Gatowski, S. I., & Dobbin, S. A. (1995). *The production of expert scientific evidence: Daubert, partisanship, and symbolic justice*. Paper presented at the Fifth International Conference on Social Justice Research, Reno, NV.
Richardson, J. T., Ginsburg, G. P., Gatowski, S. I., & Dobbin, S. A. (1995). The problem of applying *Daubert* to psychological syndrome evidence. *Judicature*, *79,* 10–17.
Risinger, D. M., & Saks, M. J. (1996). Science and non-science in the courts: *Daubert* meets handwriting identification expertise. *Iowa Law Review*, *82,* 21–74.
Saks, M.J. (2000). The aftermath of *Daubert*: An evolving jurisprudence of expert evidence. Jurimetrics, 40, 229–241.

Saxe, L., & Ben-Shakhar, G. (1999). Admissibility of polygraph tests: The application of scientific standards post-*Daubert*. *Psychology, Public Policy, and Law*, *5*(1), 203–223.
Schwartz, A. (1997). A 'dogma of empiricism' revisited: Daubert v. Merrell Dow Pharmaceuticals, Inc. and the need to resurrect the philosophical insight of Frye v. United States. *Harvard Journal of Law and Technology*, *10,* 151–237.
Schwartz, A. (2000). *The Daubert myth of the super-expert judge*. Paper presented at the meeting of the Law and Society Association, Miami, FL.
Showalter, C. R. (1995). Distinguishing science from pseudo-science in psychiatry: Expert testimony in the post-*Daubert* era. *Virginia Journal of Social Policy and Law*, *2,* 211–245.
Shuman, D. W., & Sales, B. D. (1999). The impact of *Daubert* and its progeny on the admissibility of behavioral and social science evidence. *Psychology, Public Policy, and Law*, *5*(1), 3–15.
Siegelman, P., & Donohue, J. J. (1990). Studying the iceberg from its tip: A comparison of published and unpublished employment discrimination cases. *Law and Society Review*, *24*(5), 1133–1170.
Slobogin, C. (1998). Psychiatric evidence in criminal trials: To junk or not to junk? *William and Mary Law Review*, *40,* 1–55.
Slobogin, C. (1999). The admissibility of behavioral science information in criminal trials: From primitivism to *Daubert* to voice. *Psychology, Public Policy, and Law*, *5*(1), 100–119.
Songer, D. R. (1988). Nonpublication in the United States District Courts: Official criteria versus inferences from appellate review. *Journal of Politics*, *50,* 206–215.
The American Bench (1997/1998). Sacramento, CA: Forester-Long, Inc.
United States v. Starzecpyzel, 880 F. Supp. 1027 (S.D. N.Y. 1995).
Watkins, H. (1994). Daubert v. Merrell Dow Pharmaceuticals, Inc.: General acceptance rejected. *Santa Clara Computer and Technology Law Journal*, *10,* 259–265.