

Surveying Difficult Populations: Lessons Learned from a National Survey of State Trial Court Judges*

Shirley A. Dobbin, Sophia I. Gatowski, Gerald P. Ginsburg, Mara L. Merlino, Veronica Dahir, and James T. Richardson

Surveying trial court judges presents a variety of serious challenges, but recent Supreme Court decisions and the use of increasingly complex scientific evidence in trials make it necessary to assess trial judges' knowledge and views of scientific evidence. In this article, we describe a successful nationwide survey of state trial court judges. Response rates were high, and the information obtained was valuable. Rigorous attention to detail and to proper methodology at every step is important; i.e., knowledge of the topic being investigated and the respondents' circumstances; creation and assessment of questionnaires; construction and assessment of the codebook; training and assessment of interviewers and coders; creation of analysis plans; and flexibility. This kind of care is costly and effortful, but it can make the difference between a successful survey and a waste of time and money.

There are many compelling reasons to study the judiciary. However, judges are often perceived by researchers and those who fund research as difficult to study. Reasons for this *perception* of difficulty probably include the high status and professional remoteness of the judiciary in American society, judicial time constraints, assumed resentment or unwillingness to be tested, concerns by judges about confidentiality of responses, and perhaps a distrust, dislike, or perceived irrelevance of social and behavioral science and scientists.¹ The perception that judges are difficult to study can have negative consequences

* The national survey represents one component of a larger research project: "Science in the Courtroom: Improved Understanding Through Better Judicial Education." The research project was primarily funded by the State Justice Institute (SJI#-97-02B-D3), with additional funding provided by the Federal Judicial Center; the National Judicial College; the Grant Sawyer Center for Justice Studies and Judicial Studies Program, both at the University of Nevada, Reno; and the Permanency Planning for Children Department of the National Council of Juvenile and Family Court Judges through the Office of Juvenile Justice and Delinquency Prevention, U.S. Department of Justice. This paper was originally presented at the annual meeting of the Law and Society Association, Chicago, May 1999.

¹ This perceived or actual animus of judges toward social and behavioral science research may itself derive from judges not feeling comfortable with scientific research proffered in court or from observing numerous "battles of the experts" in their courtrooms, which left triers of fact unsure how to best achieve justice. One reviewer also indicated that some judges might be reluctant to participate in research because most have to win their offices in elections, and they might fear that their responses would become public and damage their chances for reelection. This represents a special reason to be concerned that responses to surveys be kept confidential and that results are presented in ways not allowing individual identification of respondents.

for those who wish to study judges and judging. For example, funding agencies may believe that survey research projects designed to study judges cannot be conducted successfully and are, therefore, hesitant to grant funding for such studies.²

These concerns notwithstanding, many people, including social scientists, other legal professionals, and the general public, are interested in what judges do and how they do it. Indeed, given the relative power of the judiciary and the legal and social consequences of their decision making, some may argue that applied researchers have an obligation to examine judicial decision making and practices in ways that are empirically rigorous and relevant to the judges themselves. Moreover, there are some important legal and social consequences to be gained from a better understanding *by judges* of their own decision-making processes, as well as their work on the bench. Legal academics, researchers, and other observers of the legal profession should not shy away from studies of the beliefs and practices of the judiciary.

Don Dillman (1978:5) states: “[t]o conduct successful surveys, the would-be user needs a methodological ‘recipe’ that includes all the ingredients and directions for combining them.” Drawing upon our experience of conducting a recent national survey of 400 state trial court judges, we offer some insights into a “recipe” for conducting survey research with the judiciary and other populations perceived to be difficult to survey. Our recipe derives directly from Dillman’s “Total Design Method.” We will discuss the theoretical and empirical bases for making a number of key methodological decisions and offer our views on their relative success. By sharing insights we have gained from our practical experience, we hope to encourage and inspire future research efforts with the judiciary and to dispel some of the myths about judges’ unwillingness to participate in social and behavioral science research.

The Judiciary and the Legal System: Things to Consider in the Initial Planning Stages of the Research Project

Justifying Asking Potential Respondents to Participate: The research topic must be directly relevant to judges and judicial practice, the researchers must understand the issues adequately, and the topic must be presented in a reasonable fashion to potential respondents.

Members of *any* population of study should be motivated to participate to encourage their complete and timely response to survey questions. The extent that a survey includes content that will be seen as interesting, relevant, and valuable to potential respondents is an important consideration in the Total Design Method. Indeed, interest in the topic of study is a good predictor of a decision to participate in the survey project, especially when surveying busy professionals with heavy time demands. The fact that a

² This has been our experience; reviewers in several proposals we submitted to national funding agencies indicated a lack of confidence in obtaining an adequate response rate from the judiciary.

few researchers have been able to gain the participation of judges suggests that there may be particular issues that judges find important enough to respond to, methods of conducting research that judges may find more congenial, or classes of judges who may view social and behavioral science research more favorably than do others.³

The survey we discuss here, a survey of judges' understanding of and experiences with scientific evidence, occurred during a significant debate about the functioning of admissibility standards for scientific evidence in this country. A major new decision had been handed down by the United States Supreme Court that affected all federal and most state courts—the 1993 watershed decision in *Daubert v. Merrell-Dow Pharmaceuticals*, which overturned seventy years of law on the admissibility of scientific evidence.⁴ Consequently, empirical information was needed concerning the nature and extent of problems related to the use of expert evidence and testimony. The lack of empirical information on these matters was the impetus for our national survey of state trial court judges, which is, we believe, the largest such scientific survey of trial judges ever done on this topic. Moreover, our research goal was the development of a model curriculum to assist judges with understanding scientific methods and principles, uses of scientific evidence in the court, and the proper application of admissibility criteria to proffered scientific evidence—topics particularly salient to the judiciary and to judicial practice.⁵ Most judges contacted for our survey were interested in the topic, understood the value of the effort, and were willing to help, as evidenced by response rates to our “mixed-mode” survey (Dillman and Tarnai, 1988) approach, which used telephone and mail survey instruments.

Project staff spent considerable time in the early phases of the research project reviewing available legal and scientific literature regarding the admissibility of scientific evidence and implications of the *Daubert* decision. We wanted to ensure that the issues raised in the survey, and the survey questions themselves, reflected a substantial level of legal sophistication so that judges would not be deterred from participating by apparent naïveté on our part. We assumed that a knowledgeable and well-trained staff would lower error rates often obtained in studies of complex topics (Fowler, 1991).

We also pretested various versions of the instrument with focus groups of judges who were either local trial judges or visiting judges taking courses in the University of

³ A few successful major survey research projects, such as Aspin and Hall's study of retention elections and judicial behavior (1993) and Wall and Rude's study of judges as mediators (1991), have demonstrated that many judges are in fact willing to participate in survey research. Aspin and Hall, for example, achieved a response rate of .73 (N=919); Wall and Rude obtained a response rate of .71 (N=900). These response rates are not, however, the usual obtained in surveys of judges.

⁴ 113 S. Ct. 2786 (1993).

⁵ To design a curriculum that would effectively serve the needs of the judiciary, we surveyed a representative sample of 400 state trial court judges to obtain baseline information regarding judges' general scientific literacy, their understanding of admissibility criteria, and their experience in dealing with various types of scientific evidence. See Dobbin and Gatowski (1999) and www.unr.edu/bench to examine this curriculum, which contains some of the results from the survey. Other reports of specific results are available from the authors, including Gatowski et al. (2001), Dobbin et al. (1999), and Dahir et al. (2000). We also surveyed an additional “educational sample” of judges who had attended the National Judicial College to make some comparisons between that presumably more educated sample and our national random sample.

Nevada's Judicial Studies program or at the National Judicial College. This pretesting was essential for developing a survey instrument that tapped into the domain of knowledge we were studying and did so in a way that would make judges comfortable and forthcoming in their responses.⁶ (See **Appendix C** for final survey instrument, which also shows considerable detail about how the interviews were conducted.)

Accurately Framing the Target Population: Researchers must be familiar with the relevant court structures and how differences in court structures can influence the validity of any sample obtained, and they must develop an adequate sampling frame that takes into account the particular structures.

A key element in conducting *useful* research is gathering reliable and valid information. It is vitally important that the population actually surveyed is the target population of interest to the research. Care must be taken to accurately identify and sample the target population in order to draw valid conclusions from any information obtained. This may seem obvious, but studies have proceeded to unsatisfactory completion because of insufficient understanding of this aspect of the Total Design Method. Therefore, researchers wishing to study difficult populations must understand the detailed organization of the world of such populations.

In the case of judges and judging, researchers should be familiar not only with the characteristics of the target population, but also with the structure of the state and federal court systems. The structure of the federal court system is the same in all fifty states and the U.S. territories, but the structure of state court systems is established by individual state constitutions and varies from state to state. These structural differences are important to consider when designing a study that crosses state boundaries. For example, a researcher who is unaware that in the state of New York the supreme court is the trial court of general jurisdiction will find that his or her sample contains elements that should not be a part of the sampling frame if the target population is justices sitting in the highest appellate courts of every state. Likewise, in some states the district courts are the trial courts of general jurisdiction, while in other states they are courts of limited jurisdiction. Researchers unfamiliar with these distinctions may waste valuable resources before problems are identified. If such problems are not discovered before data collection begins, researchers may find that they are unable to generalize their findings to the population they intended to study.

⁶ The judges who participated in the focus groups did not like the initial structure of the survey and the wording of particular questions—they felt that they were being “tested” on scientific methods and that this was inappropriate given their professional role. The pretest subjects also felt that the survey was too long and overly complex. Despite this negative reaction, when the researchers reviewed the responses provided by the pretest subjects to the survey questions, it became clear that the questions were addressing the appropriate issues and eliciting the appropriate topic-specific responses. Thus, the pretest indicated that while the wording of the questions and the structure of the survey had to be revised, the survey itself was eliciting rich, detailed responses.

A publication that can be used effectively to construct a sampling frame for the judiciary is *The American Bench*. This biennial publication contains information about the structure of the federal and state court systems, as well as rosters of all the judges by state, organized by the level of jurisdiction within each state court system. Using information from the 1997-98 edition of *The American Bench*, we constructed a sampling frame, which consisted of all state trial court judges in the nation presiding either over courts of general jurisdiction or over courts of special jurisdiction in the states with separate divisions for juvenile and family court dockets.⁷

Our sampling frame consisted of 9,715 state trial court judges from all fifty states and the District of Columbia. A judge in the sampling frame had to be either 1) sitting on the bench of the state trial court of general jurisdiction or 2) sitting on the bench of a court of special jurisdiction, the docket of which included cases likely to contain the types of evidence relevant to the study.

Two pervasive problems in building a sampling frame are missing elements and multiple listings (Kish, 1965). Several states listed a number of vacant judicial seats at the time of publication. These missing elements were incorporated into our sampling frame by obtaining current rosters of judges from the administrative office of the courts for those states. This information was then used to supplement the initial information from *The American Bench*. Any seats listed as vacant on the current court roster were eliminated from the sampling frame. In some instances, judges were listed in *The American Bench* multiple times because they were on the bench in more than one district or jurisdiction. Duplicates were eliminated by taking the first occurrence of the judge's name and deleting all subsequent occurrences.

There were both advantages and disadvantages to using *The American Bench* as the primary source for our sampling frame. *The American Bench* offers a great deal of information in a single volume and is a useful starting place for many researchers who would like to gain a better understanding of the court system in general. From this one source we were able to ascertain 1) the structure of the state court systems; 2) the composition of the dockets within each level of jurisdiction; 3) many of the names, addresses, and telephone and fax numbers of the judges within our jurisdictions of interest; 4) biographical and demographic information about many of the judges in our final sample; and 5) information about the composition of all federal circuits.

However, a major disadvantage of using *The American Bench* is that it is published only biennially. The judiciary is not static. Judges often change offices, dockets, jurisdictions, or even careers. They are voted in and out of office, retire, receive appellate-level appointments, resign, are removed from the bench in disciplinary actions, and even die in office. Therefore, the information published in *The American Bench* becomes less reliable

⁷ We included some courts of special jurisdiction when it was clear, after confirmation with a state's administrative office of the courts, that the court docket would include cases involving types of evidence relevant to our study (i.e., social and behavioral science evidence). Excluded from the sampling frame were judges serving solely as justices of the peace because they would not be likely to hear cases involving scientific evidence.

as time passes. Investigators, if they are planning to use *The American Bench* to construct a sampling frame, must be prepared to supplement the information by referring to other directories or listings and by directly contacting courthouses or state and federal offices.

Constructing an accurate sampling frame of 9,715 names was both a difficult and a time-consuming process, taking one project staff member nearly three months to complete. To obtain reasonable levels of confidence in our findings, we had determined that we wanted a sample size of 400 for our nationwide survey.⁸ We estimated that an original sample of as many as 1,200 might be required to achieve the goal of 400 completed surveys, although we had hopes that a much smaller sample size would be required if we made serious preparation for the actual interviews.⁹ The names of all the judges in our target population were entered into a master spreadsheet in a statistical analysis program (SPSS). Judges in the sampling frame were stratified first by federal circuit and then by state. We then used proportionate stratified random sampling with a constant sampling fraction of one-eighth to obtain a sample representative of both the geographical distribution of judges and of the number of judges meeting our admission criteria in each state.¹⁰ Because we thought we might be able to achieve a high-enough response rate without using the entire sample, we then divided the total sample into two “replicates” of approximately equal size.¹¹ All judges in sample replicate A were to be contacted first. Those in sample replicate B were held in reserve to be surveyed only if the goal of 400 completed surveys was not met. Happily, this was not necessary because of our high response rate, as we were able to complete 400 interviews using only replicate A.

⁸ A sample size of 400 yields a confidence interval of about plus or minus 5 percent for most findings. We would have preferred a larger sample, but cost and time considerations precluded going beyond 400 for the main sample. Indeed, our initial grant from SJI only contained funding for a sample of 100, which we thought much too small for gathering representative information. The National Judicial College, the Grant Sawyer Center for Justice Studies, and the Federal Judicial Center supported the expansion of the main sample to 400, for which we are grateful.

⁹ See Lavrakas (1993:53-58) for how to compute this estimate.

¹⁰ By classifying elements by circuit we ensured geographic representation. Court structure and jurisdiction vary by state, so by classifying elements by state within a given federal circuit we also ensured that each state would be proportionately represented in the sample drawn. In the initial planning stages of the research we considered stratifying the sample by type of court (or docket) because types of evidence heard could vary by court jurisdiction. However, because the focus of the research was *Daubert* and the judges’ understanding of scientific methods and philosophy generally, we decided it was more important to ensure geographic representation of *Daubert* and non-*Daubert* states.

¹¹ We divided the full sample of judges obtained into two replicates using an odd-even split (i.e., judges number 1, 3, 5, etc., in sample replicate A [$n_A = 643$] and judges number 2, 4, 6, etc., in sample replicate B [$n_B = 621$]). An efficient alternative to surveying all of replicate B, had we needed to go beyond the first replicate, would be to divide Replicate B into several smaller samples, using sampling procedures with a smaller sampling fraction, thereby creating a series of smaller “subreplicates,” each equally representative of the target population. Subreplicates would be brought into play up to the point at which the desired sample size was reached. Any subreplicate brought into play would have to be implemented in full, of course. The two replicates did not contain exactly the same number because of the rule followed to assign the first judge to replicate A, the second to replicate B, and so on. Thus, when there was an odd number from a state in the sample the last one went into replicate A. Hindsight suggests that we might have made alternate or random assignments of these sample elements to equalize the size of the two replicates more closely.

Knowledgeable Staff: Project staff must be knowledgeable on issues pertinent to the survey and appropriately skilled in proper survey methodology.

When surveying difficult target populations on complex issues, it is critically important that the core staff involved in the research be experienced in survey research methodology. In our particular study it was also important that the staff was knowledgeable about the legal system generally and the issues pertaining to the admissibility of scientific evidence specifically.

Our core staff consisted of the two principal investigators, two project directors, a project coordinator, and a survey manager. The principal investigators and project directors had been part of a research team addressing issues pertaining to judicial decision making and the interface of science and law for a number of years and had produced a number of publications in this area.¹² These four individuals were responsible for all phases of project development, design, management, and analysis.¹³ All core staff were associated with the Interdisciplinary Doctoral Program in Social Psychology at the University of Nevada, Reno, either as faculty or advanced graduate students. This program requires an advanced graduate seminar in survey methodology (usually taught by Co-Principal Investigator Ginsburg) and has a major focus in the area of social psychology and law.

Fourteen interviewers were hired approximately four weeks before the first interview (graduate students in social psychology and advanced undergraduate students). Because of the high status of our population of study and the complexity of the survey instruments, we were very selective during the hiring process. Potential interviewers were referred to us by professors and project staff. Job interviews were personally conducted by the project directors, who evaluated prospective interviewers for their professionalism, interpersonal skills, ability to communicate clearly, and availability.

A comprehensive *Survey Training Manual* was developed by the project directors.¹⁴ The *Survey Training Manual* was divided into several parts. The first part addressed project-specific information: an introduction to the problem under investigation and why it is important; a statement about the purpose and goals of the project generally, and the survey specifically, and the funding sources; a discussion of the survey sample and how

¹² See for example, Dobbin and Gatowski (1998); Gatowski et al. (1997); Gatowski et al. (1996); Richardson et al. (1995); Richardson, Dobbin, and Gatowski (1995); Ginsburg and Richardson (1998); Richardson (1994, 1995, 1996, 2000); DeWitt, Richardson, and Warner (1997); Odgers and Richardson (1995); Gatowski et al. (1996).

¹³ The project coordinator and survey manager worked under the direct supervision of the project directors. The project coordinator participated in the final stages of instrument development and testing, as well as in the training and supervision of interviewers, coders, and data-entry personnel. Under the supervision of the project directors and with the assistance of the project coordinator, the survey manager oversaw the day-to-day survey operations, which included interview appointment scheduling, telephone interviewing, and interviewer scheduling, and participated in other aspects of the project.

¹⁴ The *Survey Training Manual* (Dobbin and Gatowski, 1998) developed for this project was based in part (particularly part 3) on examples of survey training manuals developed by the Survey Research Center at the University of Michigan, Ann Arbor, where one co-principal investigator (Ginsburg) had been trained (see Guenzel, Berckmans, and Cannell, 1983). Copies of the manual developed for this project are available upon request.

it was obtained; and question-by-question objectives. The second part provided specific information regarding proper survey interviewing, including the importance of question order and wording and editing of the interview. A third part dealt with how to conduct the interview, how to maintain contact, how to record responses properly, and how to use standardized probes, among other specifics. The main telephone survey instrument was included in the manual, as were copies of telephone and mail versions of a second part of the survey instrument.

All potential interviewers were required to attend eight hours of paid training conducted by the project directors. No interviewer was allowed to administer a survey without first going through the training. The first four hours of training focused on the contents of the *Training Manual*. The project directors stepped through the various components of the survey, paying particular attention to question-by-question objectives and proper interviewing techniques. The interviewers then conducted practice interviews with the project directors and the project coordinator to familiarize themselves with the survey instrument and interviewer techniques. The project directors and the project coordinator evaluated each interviewer and gave them feedback on their ability to read the questions exactly as written, their pacing and articulation, professionalism, appropriate use of standardized probes and prompts, their ability to follow skip patterns, the accuracy of their data collection, and the legibility of their handwriting. Only interviewers who successfully completed this training were allowed to begin work with the project, and more one-on-one training was given, as required. A few who completed the training were not continued on the project after assessing their initial interviews.

Achieving a High Response Rate: Initial steps involve much planning and preparation.

Interviewing difficult populations requires a great deal of preparation before making any contact with potential respondents. The initial contact must also be based on a concerted plan to obtain high response rates. Dillman (1978,1999) proposes a two-component process for conducting survey research: the Total Design Method (TDM), which focuses on the interaction between the research team and the potential respondent (also see Lavrakas, 1993). In essence, the first component of TDM calls upon the researcher to apply the tenets of social exchange theory to data collection by noting those aspects of the respondent that might influence his or her decision to participate and then taking steps to reduce the costs and increase the rewards for participation. Dillman's ideas were consistent with our own views of how to make the interview situation more valuable to the respondent and produce a higher response rate.

Letters of introduction can increase rewards and decrease costs for potential respondents. For example, a letter can reward the respondent by showing regard and appreciation. A letter can establish trust by identifying the project with a known, credible organization, or by providing a token of appreciation in advance of actual participation. A well-worded letter of introduction can also eliminate any implication of subordination on the part of the respondent by using a consulting approach (i.e., asking the respondent for help with an important area of concern). Finally, a letter of introduction can generate interest

in the research project by pointing out that the recipient should have an interest in the subject matter of the research and by showing why the subject matter is important.¹⁵

We felt that due to the high status and the possible inaccessibility of our population, the potential benefits of sending out a letter of introduction far outweighed the costs, which were mainly the time to prepare the mailing plus the postage. Our letters of introduction were personalized and printed on the letterhead of the National Judicial College (one of the survey's cosponsors) and went out over the signatures of the college's president and the director of the Master of Judicial Studies Program at the University of Nevada, Reno (one of the co-principal investigators). The letter (see **Appendix A**), which also mentioned other sponsors of the data-gathering phase of the project, outlined the nature, purpose, and goals of the research, as well as the importance of the information to be obtained, and alerted the judges that a member of the project staff would be contacting them shortly to request their participation and schedule a time to conduct an interview.

Because our sample was large and our staff was relatively small, we chose to divide the sample (replicate A) into thirds and do three mailings of the letters of introduction spaced approximately two weeks apart. By staggering the mailings, we were able to make our scheduling calls in a relatively timely manner, usually within ten to fourteen days after the letters of introduction were mailed.

During our follow-up scheduling calls, a number of judges claimed they had not yet read our letter. This was due in part to a relatively small number of out-of-date addresses in *The American Bench*. The claim was attributable, as well, to a number of other factors; for example, some judges sat in multiple districts, and in one district, although we had current correct contact information, the judge only sat in that district during certain days, weeks, or months. Clerks or secretaries screened correspondence for many judges and decided whether the correspondence was important enough to pass along. Our letter also sometimes got lost among piles of unread correspondence on a judge's desk (and the judge would locate it while the scheduler was talking to him or her). Last, but not least, in some courts the mail apparently is not distributed in a timely manner.

Sending the introductory letter in a National Judicial College (NJC) envelope worked both for us and against us. As a nationally recognized and respected judicial education organization, NJC's explicit support of the research gave our project more credibility. However, in a few cases, the envelope was discarded unopened, because the recipient assumed that it contained NJC promotional material or course listings. Nevertheless, we believe that benefits gained through the direct association and support of NJC, and the use of NJC letterhead and envelopes, far outweighed this disadvantage.

¹⁵ Merlino (1998) found that a total of seventeen of eighty-five published studies of the judiciary specifically reported sending letters of introduction before survey administration. Overall, her study revealed that the use of letters of introduction did not seem to affect response rates in the studies reviewed. Merlino notes, however, that it was difficult to estimate the impact of letters of introduction on the basis of her sample, as in many cases there was no way of determining from the article whether a letter of introduction was used. Further, there was no way of evaluating the content of the letter, nor was there any way of knowing how much time may have elapsed between receipt of the letter of introduction and the attempt to administer the survey. It was usually not clear how much follow-up there was to the letter and how the follow-up was structured. All these factors influence the utility of an introductory letter and subsequent impact on a survey's response rate.

The support of the NJC underscored the relevance of the research to the judiciary and helped to combat resistance based on assumptions that social science research would be irrelevant to day-to-day bench practice and too esoteric to be meaningful. Because the introductory letter also stated that one goal of the research was to develop a judicial curriculum (now available on www.unr.edu/bench), the explicit association and support of NJC (and the other sponsoring organizations involved in judicial education) helped, we think, in securing participation.

If the judge had not yet received or could not find our letter, the scheduler explained, if given the opportunity, the contents of the letter from a standard script, solicited participation, and then attempted to schedule the interview. Alternatively, the scheduler offered to fax a second letter and call back in the next day or two to speak with the judge again. This flexible approach was usually successful in securing a judge's participation.

Project Implementation: Develop a detailed administrative plan and choose an appropriate method, or combination of methods, for data collection.

The second component of TDM is the construction of a comprehensive and detailed administrative plan for project implementation. According to Dillman (1978:20), this administrative plan "can be viewed as essentially a decision-making enterprise constantly drawing compromises between costs and research objectives. It operates within the limitations imposed by the available responses, organization demands, and characteristics of the population under study." There are four essential aspects of an effective administrative plan: 1) all tasks to be accomplished must be identified; 2) the ways in which each task is dependent upon the performance of other tasks must be identified; 3) the order in which the tasks must be performed must be determined; and 4) how each task will be accomplished must be decided. Dillman (1978, 1999) stresses that once the administrative plan has been established, it must remain flexible enough to accommodate unexpected developments.

Like other research designs, survey research relies on a range of methodologies and procedures, which researchers in the field agree to be characteristic of well-designed and properly conducted surveys. Each type of methodology has its own advantages and disadvantages. Dillman (1978, 1999) and Frey (1989) both offer a comprehensive discussion of the strengths and weaknesses of survey research conducted by mail, telephone, and personal interview.

According to Dillman (1978, 1999), rapport and confidence building are facilitated by the face-to-face contact of personal interviews, making the personal interview the most effective way to gain the participant's cooperation. Additionally, the interviewer has greater control over the pace, flow, and order of the interview than is the case for mail or telephone surveys. The trade-off for the rapport and control of the personal interview is the time and expense involved with data collection. Personal interviews require highly trained interviewers who are in the same geographical location as the respondents. For this reason alone, the time and expense of conducting personal interviews on a geographically diverse population is usually prohibitive.

In an archival study of survey methodology used in social and behavioral science research in which judges were the participants, Merlino (1998) found that only eleven of the eighty-five published studies included in her sample chose personal interviewing as the primary method for data collection. The average response rate reported for these studies was 80 percent ($n=11$, $M=.80$, $SD=.25$). While this overall response rate is good, it is important to note that the number of respondents ranged from 3 to 264, and the next greatest number of respondents below the maximum of 264 was 36. The median number of respondents was 7.

Dillman (1978, 1999), Frey (1989), Lavrakas (1993), and especially Groves (1989:533-39) all argue that telephone surveys are generally less expensive to conduct than personal interviews and are particularly appropriate if the target population is large or geographically diverse, as was the case with our survey. Interviews can be conducted from a central location, which facilitates interviewer staffing and supervision, as well as interviewer safety. Dillman (1978) reports that the response rates for telephone surveys are somewhat lower than those for personal interviews, however, and members of the target population may actually be omitted from the sample because they do not have telephones (not a problem when interviewing judges or other high-status populations, of course).

Merlino (1998) found that only two of the eighty-five published studies in her sample used telephone interviews, a somewhat surprising finding. The average response rate for these two studies was 71 percent ($n=2$, $M=.71$, $SD=.35$), although the two studies varied greatly. In one, the response rate for $n=31$ was 95 percent, while in the other, the response rate obtained for $n=96$ was 46 percent. The first study used a combination of sampling techniques, while the latter used a nonrandom sampling design with little systematic follow-up.

Mail surveys are the least costly to conduct (Dillman, 1978, 1999). The time and expense involved in training and supervising interviewers is not a factor, nor is geographical location. Response rates for mail surveys are typically much lower than those for the other methodologies, however. Out-of-date or incomplete address listings often result in surveys being returned as undeliverable. Lack of response may also be attributed to inadequate reading or writing skills of the respondent or even to the design of the questionnaire itself (Dillman, 1978; Sanchez, 1992).

Merlino's (1998) findings were consistent with Dillman's (1978) with respect to the generally low response rates associated with mail surveys. She found that the response rates for the mail surveys in her sample ranged from .05 ($n=103$) to 1.00 ($n=5$ and $n=35$) and averaged 48 percent ($n=51$, $M=.48$, $SD=.23$). Despite this relatively low average response rate, 51 of the 85 (60 percent) surveys in her sample were conducted by mail.

When deciding which survey methodology would best suit our goals, we considered the size and geographic location of our intended sample, the cost, the issues addressed by the survey, the complexity and length of the survey, inclusion of and the amount of both open- and close-ended questions, feedback obtained from several pilot tests of the survey instrument, and suggestions from our advisory committee.¹⁶ Our final survey

¹⁶ Our major funding agency, the State Justice Institute, required a national advisory committee to review all aspects of the project, a group that served us well.

design combined telephone and mail methodologies. The first part of the survey, which gathered opinions about the utility of the *Daubert* guidelines and views on additional issues, such as the use of expert witnesses and jurors' comprehension of expert testimony, was administered via telephone. Judges could complete the second component of the survey, which consisted primarily of closed-ended questions and rating scales, either by mail or by telephone.¹⁷ This decision enabled us to minimize the compromise between data quantity and quality—we maintained the high quality of our data by using the telephone for most of the open-ended questions, while increasing the quantity of data by using a shorter mail instrument for closed-ended items. We also anticipated a high response rate for the mail surveys given the amount of time the participants had already invested in responding to the first part of the survey by telephone, and this expectation was borne out. We had an overall response rate on part 2 of 81 percent (325 of 400). Of the 400 judges surveyed, 123 (31 percent) chose to complete part 2 via telephone immediately after completion of part 1, and 277 judges (69 percent) chose to complete part 2 via mail. For the subsample of judges who received part 2 in the mail, we had a return rate of 73 percent (207 surveys returned of a possible 277).

The Judiciary and the Legal System: Practical Survey Management in the Real World

Scheduling Interviews

First, a survey-tracking database of all 643 judges from replicate A was developed. The database listed the judge by name, his or her bench, contact information, and corresponding six-digit numerical code identifier assigned to protect the judge's identity. The database was designed to allow the project staff to track the scheduling and rescheduling of interviews, the date of completion of telephone interviews, and whether the judge had completed only part 1 of the interview or both part 1 and part 2. For those judges completing only part 1 by phone, the database also tracked the dates in which part 2 was mailed and returned, and any necessary follow-up contacts. This database was updated at the end of each day, which was essential to the organization and management of the data.

When following up on the letters of introduction, schedulers used a standardized script to explain further the importance of the research and to encourage the judges to participate. This script provided schedulers with responses to anticipated questions or concerns from judges (e.g., confidentiality, use of case-specific information) and standard messages to leave on answering machines or with secretaries or court clerks. All contacts and attempted contacts, including the name of the judge's clerk or secretary, were logged on the judge's contact sheet, or "call sheet," with the date, time, the initials of the persons attempting the contact, and any relevant comments. The disposition for

¹⁷ This strategy has been used successfully by the Federal Judicial Center in surveys of federal judges and was suggested by Joe Cecil, a member of our advisory committee and director of the Research Division at the Federal Judicial Center.

each call (e.g., if the scheduler received a busy signal, spoke to the secretary, or left a message on the judge's voice mail) was also recorded. On many occasions, the scheduler was told to call back on a specific date and time. This date and time for a callback was noted on the call sheet and the call sheet was filed in a "tickler system" by date to ensure that callbacks were conducted as promised. All call sheets for interviews to be conducted were also filed under the tickler system to ensure that judges were scheduled on the proper date and at the proper time (keeping close track of time zones was essential). Interview packets were assembled using these call sheets as a guide (see **Appendix B** for a sample call sheet).

If the judge agreed to be interviewed, the scheduler made an appointment for an interviewer to call at the judge's convenience. If the judge declined to be interviewed, the scheduler asked the judge for his or her reasons for refusing and noted the judge's response on the call sheet. The scheduler was also asked to evaluate the strength of the judge's refusal and to indicate whether he or she thought a project director should follow-up with the particular judge in the hopes of securing participation. This was done using a seven-level Likert scale for three questions ("strength of refusal," from very weak to very strong; "respondent attitude," from very polite to very rude; and "level of anger," from not at all angry to very angry), as well as an open-ended statement of reason given for refusal and an assessment by the interviewer of whether the person should be called back by someone else. Recording reasons for declining and the strength of refusal enabled the survey staff to identify those judges who declined due to time constraints or other temporary circumstances and to contact them at a later date to see whether their circumstances might then allow them to participate at that time. This information saved time because it allowed project staff to readily identify those respondents who would probably never agree to an interview. The calling of refusals was done infrequently, but did succeed in gaining a few more completed interviews, thus raising the response rate.

Completing Interviews

We did not establish a maximum number of scheduling calls to be made before the judge was considered a refusal. Rather, recognizing that it might be difficult to reach judges, and that we might have to leave multiple messages and engage in callbacks, schedulers were instructed to continue to contact judges (e.g., leave messages and play "phone tag") until the judges agreed to participate or gave firm refusals. Thus, we chose to keep making scheduling calls until the end of our grant period to achieve as high a response rate as possible. Three people handled all callbacks and maintained systematic records of callbacks during the project.

The total number of calls made to complete the 400 interviews was 1,603, which does not include, of course, the calls made to judges who were never interviewed. On average, it took four calls to make contact with most judges (range of one to seventeen calls). Approximately 40 percent of the judges (n=162) were contacted within two calls, and approximately two-thirds were contacted within four calls (n=268). Ten percent of the judges, however, required more than eight calls before direct contact was established.

Beyond twelve calls, the yield per call was less than 1 percent. These data suggest that even fewer than twelve callbacks might be justified, even with hard-to-survey populations.

Krosnick (1999:540) makes the point that very difficult to reach respondents may not be representative of the greater population, which also argues against unlimited callbacks. One might argue, however, that being “difficult to contact” is a professional characteristic of not only judges but also other populations. It is important to recognize that while we were able to contact two-thirds of the judges interviewed within four calls, this completion rate was probably facilitated by the level of preparation steps taken to secure participation and the credibility and professional relevance derived through association with our funding partners.

Fifty-nine of the 643 judges (9.2 percent) constituting replicate A could not be reached at all by the survey staff. Of these 59 judges, 53 (89.8 percent) did not return phone calls or respond to messages. The administrative assistants of another 3 judges (5 percent) strongly suggested that the judges were probably not interested in participating, but the judges themselves did not specifically decline to be interviewed. The remaining 3 judges were not contacted because of an oversight not noticed until the project was completed (the call sheets were misfiled).

One hundred and twelve of the 643 judges (17.4 percent) explicitly declined to participate.¹⁸ Of these 112 judges who explicitly declined to participate, 49 (43.8 percent) indicated that time constraints of some sort made participation inconvenient. Sixteen judges who refused (14.3 percent of 112) stated that they lacked relevant experience, either because they were new to the bench or because they did not have experience with scientific evidence. Five refusing judges (4.5 percent of 112) stated that they were retired and not doing trials. Twenty-four judges (21.4 percent of 112) stated that they were “just not interested” in participating without giving any further reason, 2 (1.8 percent) said they could not for “confidentiality reasons,” and the rest (14.3 percent) declined to state a reason. We also had 42 judges who withdrew after the interview had started or after they had committed to be interviewed, mainly because of claimed time constraints or because they said they believed they lacked experience enough to respond properly. Thus, we had 400 completed, usable interviews from the sample, representing a response rate of 71 percent on part 1.¹⁹ The breakdown is as follows:

¹⁸ We encountered some hesitancy about agreeing to be surveyed in one particular state where judges had recently been asked to participate in a survey with a payment promised for participation. Such payments may violate judicial canons of ethics and should not be used with judges. Either a project director or project coordinator spoke with these judges to assure them that compensation was not being offered and that participation in the survey did not violate their canon of ethics.

¹⁹ Ten completed interviews were uncodeable and are not included in the final sample of 400. The survey literature disagrees about the calculation of response rates. One way is to include in the denominator only those eligible target persons actually contacted, thereby revealing the rate of actual refusals. The response rate of .71 was calculated based on 400 completed interviews of the 564 eligible persons actually contacted (from 643 persons in replicate A subtract 20 out-of-frame elements [retired, had died, no longer on bench], and subtract 59 eligible elements never contacted = 564). A more conservative approach is to include all listings even if never contacted, but that overestimates the refusal rate and the potential for bias. Following the second approach, the response rate for part 1 would be .64 (400 of 623 eligible elements; the 59 no-contacts are included in the eligible elements).

Total in replicate A	643
Out-of-frame elements	20
No contact	59
Explicit refusals	112
Withdrawals	42
Uncodeable interviews	<u>10</u>
Total usable interviews	400

The telephone component averaged about one hour to complete, with a range of from 20 minutes to 150 minutes. Considering the substantial time commitment involved, most of the judges in our sample were very cordial and generous with their time, but we found that we had to be extremely flexible to accommodate their schedules. When making scheduling calls, we were candid about the time commitment involved, and always scheduled appointments at the judge's convenience. This meant that we had to be prepared to conduct interviews at all hours of the day or night, including weekends and holidays. And because we were drawing from a national sample, we had to accommodate interviews in time zones from Hawaii to Maine. Our earliest morning interview was conducted at 4:00 a.m., and the latest evening interview at 10:00 p.m. Most of the interviews occurred between 8:00 a.m. and 3:00 p.m. (Pacific Time).

We often found that the judge was still on the bench or in a meeting at the scheduled interview time. If this happened, we either had the interviewer continue to try to reach the judge in the thirty minutes following the scheduled interview time or reschedule the appointment for another date and time with the judge's clerk or secretary. In several instances, judges were called away in the middle of the interview. In this event, we would schedule a date and time to complete the interview and, whenever possible, assign the original interviewer to finish conducting the interview. Accommodating the judge's schedule not only meant being prepared to push back appointment times or reschedule interviews, but also meant being prepared to conduct interviews without notice when a judge told the scheduler that he or she was available right at that moment, a practice we came to refer to as "scrambling for an interviewer." We handled this situation by arranging to call the judge back in ten minutes. During that ten minutes the scheduler notified the supervisor that the judge wanted to do the interview immediately, then assembled an interview packet. The supervisor would then check the availability of any of our trained interviewers who happened to be available. If an interviewer was available, then the supervisor would assign that person a phone line and the interviewer would call the judge back. If no interviewer was available, then either the scheduler (who had also been cross-trained as an interviewer), the project coordinator, the survey manager, or a project director would conduct the interview.

Survey Management

Once an interview was scheduled, the survey manager or project coordinator posted the date and time of the interview and the name of the judge on a master scheduling board located in the central office for the survey team. A large "white board" was used

so that changes in scheduling times and interviewers could be easily made and were easy to read. At the end of each day, interview packets for the next day's interviews were compiled for all scheduled interviews. Interview packets contained 1) the judge's call sheet and appropriate phone number to call for the interview; 2) telephone survey instruments for both the first and second part of the survey; 3) a sheet with a list of acceptable probes and feedback phrases; 4) a printed postcard, signed by one of the project directors, thanking the judge for participating in the project, which the interviewer signed and mailed after the interview; 5) a sheet containing case citations and a standardized explanation of the *Daubert* decision; and 6) several sharp, number-two lead pencils.

Interviewers were required to meet the supervisor at the interview facility fifteen minutes before the scheduled interview time to pick up the interview packet and to receive their office assignments. Once the interview began, the supervisor periodically looked in on the interviewer to make certain that the interviewer had everything he or she needed, was adhering to professional interviewing techniques, and had no problems or questions that needed to be resolved.²⁰ Initially, the interviews were supervised quite heavily. As the interviewers gained more experience, however, they ultimately required minimal supervision. After completing an interview, the interviewer signed the thank-you card and returned the packet to the supervisor. If the judge completed only the first portion of the survey over the phone (recall that the respondents were given the option of completing both parts of the survey by phone or completing the second part by mail), the supervisor then prepared an envelope containing the second portion of the survey, the thank-you card, and a self-addressed, stamped return envelope. These envelopes were mailed out each day after the final interview had been completed. If the judge was not available for the interview, the interviewer returned the packet to the supervisor and reported to the scheduling facility to work as needed on scheduling, coding, or data entry.

Several of our most competent interviewers were offered additional work coding, scheduling, entering data, and conducting literature searches. We benefited from this in several ways. For example, we paid our interviewers for a minimum of half an hour if their scheduled interviews fell through. By cross-training our interviewers, we facilitated the development of our codebooks, which were empirically constructed using the responses from the first hundred surveys, and kept the coding and data entry going concurrently with the interviewing. This both reduced the workload for the core project staff and allowed our best interviewers to earn enough money to justify their remaining with the project until its conclusion.

Coding of Responses

An empirical code book was developed for both the close-ended and the open-ended questions on the survey instruments. Open-ended codes were developed based upon a

²⁰ We did not use a centrally located phone bank for reasons of cost, but used phones made available to us in offices in two adjoining buildings housing the Grant Sawyer Center for Justice Studies and the National Judicial College. This complicated project oversight and supervision of interviews, but we resolved those difficulties by having supervisors moving between offices in which interviews were taking place and by using a cell phone so the supervisors could be contacted immediately if any question or problem arose during an interview.

random drawing of 50 of the first 100 completed interviews. Responses to open-ended questions were reviewed, and mutually exclusive and exhaustive codes were generated 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 received a day of training on the project, including familiarization with the objectives of the study and of each question, along with practice coding on interviews, with corrective feedback.

As previously mentioned, some of our most competent interviewers were cross-trained for coding and data entry. Cross-training not only maximized limited resources, but also meant that those individuals coding and entering data had an in-depth understanding of the survey instrument and the interview process. And because their level of competency had been recognized by the core project staff and additional responsibilities had been assigned, our coders were committed to the project and to high-quality, accurate coding and data entry.

Approximately 25 percent of the surveys coded were check-coded (Cohen kappa = .84) and showed high, chance-corrected reliability. Any disagreements that arose during check coding were resolved, and if they required minor changes in the code book, those changes were made and conveyed to all coders. Any retroactive impact of changes was also implemented. Coding was recorded on code sheets to facilitate check coding, and then entered into SPSS for data analysis, with 100 percent of the code sheets checked for data entry accuracy and any discrepancies corrected.

Conclusions

As discussed earlier, Dillman's (1978, 1999) Total Design Method calls for the construction of a comprehensive and detailed administrative plan for project implementation. The administrative plan consists of four components: 1) all tasks to be accomplished must be identified; 2) the ways in which each task is dependent on the performance of other tasks must be identified; 3) the order in which the tasks must be performed must be determined; and 4) how each task will be accomplished must be decided. Dillman also stresses that the administrative plan must remain flexible enough to accommodate the unexpected, which will inevitably arise.

Using this very organized method of administration, coupled with meticulous attention to detail, strict adherence to the rules of sampling, and a personalized and flexible approach to gaining cooperation from respondents, we were able to successfully complete this major survey of judges on a complex topic, the use of science in the courtroom. Our methods allowed us not only to survey a difficult population (judges), but to do so in a rigorous way that garnered high-quality data. We hope others will be encouraged by this success and be guided by our attention to detail.

On the basis of our experience surveying judges, we cannot stress enough the importance of flexibility. We did our best to identify all tasks and to establish procedures for accomplishing them. However, the tasks related to managing a large-scale survey project can be likened in some ways to dominos. When one domino is added, shifted, or

removed, the way the other dominos fall is also changed. Similarly, changes in survey management procedures affect the way that subsequent tasks are performed. Flexibility enabled our project staff to arrive at creative solutions to unanticipated events, such as faxing letters of introduction to judges who for some reason did not receive the original letter or establishing communication with the survey supervisor by cellular phone to capitalize on fortuitous breaks in our judges' schedules.

It is important to note, however, that we did not alter our rigorous standards for high-quality interviewing and data gathering through our flexible approach. Rather, by creating new procedures and changing or eliminating practices that were impractical or ineffective, we were able to increase our response rate and gather data of the highest quality by ensuring that no judge who was available for an interview was lost due to unavailability of staff and that sufficient resources and high-quality staff were available throughout the data collection, coding, and data entry phases of the project. Thus, we achieved good response rates from a project that was complex and involved lengthy mixed-mode instrumentation, proving that difficult-to-survey populations can be studied using properly applied survey methods. **jsj**

REFERENCES

- The American Bench: Judges of the Nation*, 9th ed. (1997-98). Sacramento, CA: Forster Long, Inc.
- ASPIN, L. T., and W. K. HALL (1994). "Retention Elections and Judicial Behavior," 77 *Judicature* 306.
- DAHIR, V., M. L. MERLINO, S. A. DOBBIN, S. I. GATOWSKI, J. T. RICHARDSON, and G. P. GINSBURG (1999). "A Clash of Cultures: When Diverse Disciplines Meet in the Legal Arena." Paper presented at the annual meeting of the Law and Society Association, Chicago.
- DEWITT, J. S., J. T. RICHARDSON, and L. W. WARNER (1997). "Novel Scientific Evidence in Controversial Cases: A Social Psychological Analysis," 21 *Law and Psychology Review* 1.
- DILLMAN, D. A. (1999). *Mail and Internet Surveys: The Tailored Design Method*. New York: John Wiley and Sons.
- (1978). *Mail and Telephone Surveys: The Total Design Method*. New York: John Wiley and Sons.

- DILLMAN, D. A., and J. TARNAI (1988). "Administrative Issues in Mixed Mode Surveys," in R. M. Groves et al. (eds.), *Telephone Survey Methodology*. New York: John Wiley and Sons.
- DOBBIN, S. A., and S. I. GATOWSKI (1999). *A Judge's Benchbook on the Basic Philosophies and Methods of Science*. Washington, DC: State Justice Institute.
- (1998). "The Social Production of Rape Trauma Syndrome as Science and Evidence," in M. Freedman and H. Reece (eds.), *Science in Court*. London: Dartmouth Press.
- (1998). *Survey Training Manual for "Science in the Courtroom: Improved Understanding Through Better Judicial Education."* Reno, NV: Grant Sawyer Center for Justice Studies.
- DOBBIN, S. A., S. I. GATOWSKI, J. T. RICHARDSON, G. P. GINSBURG, M. L. MERLINO, and V. DAHIR (1999). "Judging Science: Results from a National Survey of Trial Judges about *Daubert* and Scientific Evidence." Paper presented at the annual meeting of the American Psychology-Law Society, Dublin.
- FOWLER, F. J. (1991). "Reducing Interviewer-related Error Through Interviewer Training, Supervision and Other Means," in P. R. Biemer et al. (eds.), *Measurement Errors in Surveys*. New York: John Wiley and Sons.
- FREY, J. (1989). *Survey Research by Telephone*, 2nd ed. Newbury Park, CA: Sage.
- GATOWSKI, S. I., S. A. DOBBIN, J. T. RICHARDSON, G. P. GINSBURG, M. L. MERLINO, and V. DAHIR (2001). "Judging Science in a Post-*Daubert* World: Results of a National Survey of Judges," 25:5 *Law and Human Behavior* 433.
- GATOWSKI, S. I., S. A. DOBBIN, J. T. RICHARDSON, and G. P. GINSBURG (1997). "The Globalization of Behavioral Science Evidence About Battered Women," 15 *Behavioral Sciences and the Law* 285.
- GATOWSKI, S. I., S. A. DOBBIN, J. S. DEWITT, J. F. CARPENTER, R. C. DILLEHAY, J. T. RICHARDSON, and D. CARNES (1996). "Court-annexed Arbitration in Clark County, Nevada: An Evaluation of Its Impact on the Pace, Cost, and Quality of Civil Justice," 18 *Justice System Journal* 287.
- GATOWSKI, S. I., S. A. DOBBIN, J. T. RICHARDSON, C. NOWLIN, and G. P. GINSBURG (1996). "The Diffusion of Scientific Evidence: A Comparative Analysis of Admissibility Standards in Australia, Canada, England, and the United States, and Their Impact on the Social and Behavioral Sciences," 4 *Expert Evidence* 86.

- GINSBURG, G. P., and J. T. RICHARDSON (1998). "'Brainwashing' Evidence in Light of *Daubert*," in H. Reece (ed.), *Law and Science*. Oxford: Oxford University Press.
- GROVES, R. M. (1989). *Survey Errors and Survey Costs*. New York: John Wiley and Sons.
- GUENZEL, P. J., T. R. BERCKMANS, and C. F. CANNELL (1983). *General Interviewing Techniques: A Self-instructional Workbook for Telephone and Personal Interviewer Training*. Ann Arbor: Survey Research Center, Institute for Social Research, University of Michigan.
- KISH, L. (1965). *Survey Sampling*. New York: John Wiley and Sons.
- KROSNICK, J. A. (1999). "Survey Research," in J. T. Spence, J. M. Darley, and D. J. Foss (eds.), *50 Annual Review of Psychology* 537.
- LAVRAKAS, P. J. (1993). *Telephone Survey Methods: Sampling, Selection, and Supervision*. Newbury Park, CA: Sage.
- MERLINO, M. L. (1998). "Judicial Participation in Social and Behavioral Science Research." Paper presented at annual meeting of Law and Society Association, Aspen, Colorado.
- ODGERS, S. J., and J. T. RICHARDSON (1995). "Keeping Bad Science out of the Courtroom: Changes in American and Australian Expert Evidence Law," 18 *University of New South Wales Law Journal* 108.
- RICHARDSON, J. T. (2000). "Discretion and Discrimination: Problematic Evidence in Cases Involving Controversial Religious Groups and Accusations of Satanic Ritual Abuse," in R. Adhar (ed.), *Law and Science*. Burlington, VT: Ashgate.
- (1996). "'Brainwashing' Claims and Minority Religions Outside the United States: Cultural Diffusion of a Questionable Concept in the Legal Arena," *Brigham Young University Law Review* 873.
- (1995). "Minority Religions ('Cults') and the Law: Comparisons of the United States, Europe, and Australia," 18 *University of Queensland Law Journal* 182.
- (1994). "Dramatic Changes in American Expert Evidence Law: From *Frye* to *Daubert*," 2 *Judicial Review* 13.

- RICHARDSON, J. T., G. P. GINSBURG, S. I. GATOWSKI, and S. A. DOBBIN (1995). "Problems Applying *Daubert* to Psychological Syndrome Evidence," 79 *Judicature* 17.
- RICHARDSON, J. T., S. A. DOBBIN, and S. I. GATOWSKI (1995). "The *Daubert* Decision Concerning Scientific Evidence," 3 *Nevada Lawyer* 13.
- SANCHEZ, M. E. (1992). "Effects of Questionnaire Design on the Quality of Survey Data," 56 *Public Opinion Quarterly* 206.
- WALL, J. A., and D. E. RUDE (1991). "The Judge as Mediator," 67 *Journal of Applied Psychology* 54.

APPENDIX A

Letter of Introduction (on National Judicial College Letterhead)

Dear Judge_____:

The 1993 *Daubert* decision (113 S. Ct. 2786, 1993), which may revolutionize the admissibility of scientific evidence in federal, as well as many state courts, is but one major sign that courts at every level are dealing with scientific evidence of increasing complexity. *Daubert* and other decisions have increased the gatekeeping responsibilities of trial judges who, as a result, are being required to understand the nature of science and the scientific method more thoroughly.

Recognizing that current judicial education programs may not provide judges with sufficient training to meet these new responsibilities, the State Justice Institute has generously awarded the Grant Sawyer Center for Justice Studies and the Master of Judicial Studies Program at the University of Nevada, Reno, a research grant to assess the specific needs of the judiciary and to develop a curriculum which will be used to facilitate judges' understanding of science in the courtroom. The National Judicial College, the Federal Judicial Center, and the National Council of Juvenile and Family Court Judges have also contributed considerable support to ensure the success of this research project.

We are contacting you to ask for your help in meeting our objectives in this important nation-wide study. Your participation will involve a telephone interview concerning admissibility standards and scientific evidence. This interview will require approximately 45 minutes to conduct, and will be scheduled at your convenience.

Of the approximately ten thousand state trial court judges eligible to be part of this research effort, you have been selected to participate in this study. Your cooperation is extremely important to ensure that the results are truly representative of both your needs and the needs of your colleagues. One of the research assistants on this project will be contacting you shortly by telephone to see if you are willing to participate in this research and to schedule a convenient time for your telephone interview. On behalf of the State Justice Institute, The National Judicial College, the Federal Judicial Center, and the National Council of Juvenile and Family Court Judges, we thank you in advance for this service to us and to your colleagues.

Respectfully,

Judge Robert Payant
President of The National Judicial
College

James T. Richardson, J.D., Ph.D.
Director, Master of Judicial Studies
Program

APPENDIX B

CALL SHEET

CONTACT:

TELEPHONE NUMBER:

Scheduled Interview

DATE: _____ **TIME:** _____ (Interviewee Time Zone)
 _____ PACIFIC STD. TIME

Secretary/Office Contact: _____

Contact	Date	Time	Disposition Code/Notes	Initials
1.	___/98	_____	_____	_____
2.	___/98	_____	_____	_____
3.	___/98	_____	_____	_____
4.	___/98	_____	_____	_____
5.	___/98	_____	_____	_____
6.	___/98	_____	_____	_____
7.	___/98	_____	_____	_____
8.	___/98	_____	_____	_____

Notes

1. _____
2. _____
3. _____
4. _____
5. _____
6. _____
7. _____
8. _____

Refusals

Strength of Refusal: VERY WEAK 1—2—3—4—5—6—7 VERY STRONG

Respondent Attitude: VERY POLITE 1—2—3—4—5—6—7 VERY RUDE

NOT AT ALL ANGRY 1—2—3—4—5—6—7 VERY ANGRY

Reasons given for not wanting to participate: _____

Do you think the individual should be called back by someone else to try to persuade him/her to participate? YES NO (Circle one)

Suggestions: _____

APPENDIX C

Example Segments from Mail Questionnaire (©1998)

Following are a few selected questions from the large survey instruments developed for this research. The complete questionnaires are available at www.unr.edu/bench along with the code books we used, which are essential to understanding how we actually coded the information and made use of it in our research reports.

The first section pertains to specific forms of expert evidence that may be encountered in the courtroom. For each of the specific types of evidence I'm going to list, I would like you to rate the amount of courtroom experience you have had with that evidence on a scale of 0 (no experience at all) to 5 (a great deal of experience).

- 1A. The first type of evidence is epidemiological evidence. That is, evidence related to the study of diseases, their distributions in populations, and their environmental evidence. On a scale from 0 (no experience at all) to 5 (a great deal of experience), how much courtroom experience have you had with epidemiological evidence? [CIRCLE APPROPRIATE NUMBER]

0————1————2————3————4————5
 No Experience Great Deal
of Experience

- 1B. What about DNA evidence, which might be, for example, proffered as evidence of a match between a sample obtained from a crime scene and DNA extracted from an individual? How much courtroom experience have you had with DNA evidence? [CIRCLE APPROPRIATE NUMBER]

0————1————2————3————4————5
 No Experience Great Deal
of Experience

- 1C. What about psychological syndrome evidence? Evidence related to a psychological syndrome is generally defined as a group of behavioral or psychological characteristics that appear to occur together and to be interrelated. On the scale from 0 to 5, how much courtroom experience have you had with psychological syndrome evidence? [CIRCLE APPROPRIATE NUMBER]

0————1————2————3————4————5
 No Experience Great Deal
of Experience

1D. And lastly, what about psychological profile evidence? A psychological profile is generally defined as a set of symptoms, characteristics, or behaviors that are associated with specific behavior patterns. How much courtroom experience have you had with psychological profile evidence? [CIRCLE APPROPRIATE NUMBER]



PSYCHOLOGICAL SYNDROME EVIDENCE

IF RESPONDENT INDICATED NO EXPERIENCE WITH PSYCHOLOGICAL SYNDROME EVIDENCE (RESPONDED WITH ‘O’ IN Q#1C), THEN SKIP TO P.9, Q#8. OTHERWISE, ASK Q#4.

Now, please consider psychological syndrome evidence. I’m going to list different psychological syndromes and I would like you to rate, on the same scale of 0 (no experience at all) to 5 (a great deal of experience), the amount of courtroom experience you have had with each syndrome.

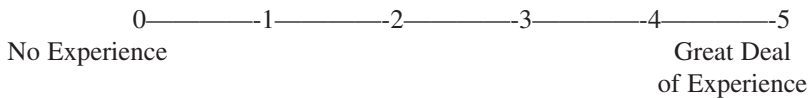
4A. First, battered woman syndrome. On a scale of 0 (no experience at all) to 5 (a great deal of experience), how much courtroom experience have you had with battered woman syndrome?



4B. What about rape trauma syndrome?



4C. What about child sex abuse accommodation syndrome



4D. And what about parental alienation syndrome?



4E. And, lastly, what about repressed memory syndrome? On a scale of 0 (no experience at all) to 5 (a great deal of experience), how much courtroom experience have you had with repressed memory syndrome?



5. Now, on the same scale of 0 to 5, how much courtroom experience have you had with post-traumatic stress disorder?



6. Other than those already mentioned, have you had experience with any other psychological syndrome evidence?

No (**SKIP TO QUESTION #7**)

Yes

6A. What were they?

Thinking now of psychological syndromes in general

7. Approximately how many cases have you had where the admissibility of psychological syndrome evidence was at issue?

7A. In your experience, what has been the most common type of case in which psychological syndrome evidence is proffered?

7B. For those cases in which the admissibility of psychological syndrome evidence was at issue, in approximately what percentage of cases was the psychological syndrome evidence admitted? _____%

7C. And were all the rest rejected?

[] Yes

[] No

7D. What factors do you consider when determining the admissibility of psychological syndrome evidence?

7E. In your experience, what aspect of psychological syndrome evidence is most problematic in determining its admissibility?

7F. In terms of admissibility, are there any psychological syndromes that have come before your court which have not been problematic?

[] No (**SKIP TO QUESTION #7G**)

[] Yes

7Fi. Which ones?

7Fii. And why has that particular psychological syndrome, or syndromes, not proven to be problematic for the court?
