

Judicial Performance Evaluation

Moderator

Noel Brennan

Project Director, American Bar Association Judicial Performance Evaluation Guidelines Project.

Panel Presenters

Francis Bremson

Executive Director, Alaska Judicial Council.

The Honorable Alan Handler

Justice, Supreme Court of New Jersey.

Janet Franz

Assistant Trial Court Administrator, Supreme Court of Vermont.

Noel Brennan

In July of 1985 the American Bar Association adopted, as policy, guidelines for the evaluation of judicial performance. The history of the guidelines project is somewhat long, with the idea originating in 1979. A committee of the American Bar Association considered whether to undertake a project that attempted to rationalize the thinking and the emotional pitch of judicial performance evaluation. A concept paper was developed and funding sought and secured, and in November of 1983 a special committee and two drafting task forces, for which I served as the staff director, were formed. We were fortunate in having among the resources of the task forces Justice Alan Handler from the New Jersey Supreme Court who helped us think through the Association's policy in an area in which we are all still learning.

The subject of judicial evaluation is not a popular one. The judicial branch has traditionally been reluctant to become involved in assessment or evaluation of its judges. Unsatisfactory experiences with bar polls and special interest groups have made judges rightfully wary of any program of judicial evaluation.

This workshop presents three panelists on the subject of reliable performance evaluation programs which have been developed and refined. Alaska has a full scale process which focuses on retention and includes the public dissemination of information. New Jersey has an internally administered program whose primary

goal is the improvement of judges' performance rather than the facilitation of selection or retention processes. Similarly, Vermont recently initiated a judicially sponsored program aimed at personal assessment and feedback for the sitting judge.

To help set the stage for the panelists, I am going to ask each of them in the course of describing their programs to comment on the following questions:

- 1) What are the major areas of judicial performance tested by your program?
- 2) In your experience, is it feasible to obtain reliable data?
- 3) Can results be presented in a form that is useful to a judge?
- 4) Have procedures been adopted for anonymity and confidentiality?
- 5) Have you heard complaints from individual judges about impaired judicial independence or integrity as a result of the initiation of your program?
- 6) Have you taken any steps in your program to insure that respondent data is free from gender bias?

Francis Bremson

I am the staff director of the Alaska Judicial Council, which is an oversight body that performs evaluations for the purpose of retaining judges. I will address how judicial evaluation started in Alaska and the judges' reactions to evaluation.

First of all, the Constitution of Alaska, which was adopted in 1955 but became effective in 1959 when Alaska became a state, provided for retention elections. The term "retention elections" means that when judges' terms are over, they run for reelection on their record. When judges were first appointed in Alaska, they knew at the end of their term they would be eligible to run for retention on their record. By and large, they ran with little or no fanfare and with little or no evaluation data available to the public. From time to time the Bar Association polled its members regarding whether judges standing for retention were qualified or unqualified, and the results of these polls were circulated generally to the members of the Bar. Nevertheless, those polls did not have a great deal of impact on retention.

In the early 1960's, a conflict between the bench and bar resulted in a campaign by a faction of the Bar to remove a supreme court justice standing for retention. The justice was defeated and shortly thereafter Alaska had an integrated bar association.

In the late 1960's or early 1970's, the Bar Association began

conducting evaluation polls, using the "Dallas model" (apparently taken from a poll conducted by the Dallas Bar Association). At about the same time, the Judicial Council decided it ought to do an evaluation of judges standing for retention and to make those results public. Responsibility for conducting the evaluation was debated back and forth, with the Judicial Council and the Bar conducting separate polls in 1974.

In 1975, the issue was kicked over to the Legislature. The Legislature, in effect, said, "Let the Judicial Council conduct these evaluations and be responsible for conducting these polls." The Bar Association was involved in that legislative process and agreed to let the Judicial Council conduct the polls. The lieutenant governor, who by statute was required to provide information regarding all general elections to the public, also got into the act by agreeing to include judicial performance evaluation information in the voters' pamphlets, which were (and still are) provided to every voter in the state.

The first poll of members of the Bar Association regarding the judges standing for retention took place in 1976. That year there was also a poll taken of all law enforcement officers in the state. Critics of the police survey charged that such a survey accorded more influence to a special interest group than was justified. While polling of lawyers is easily supported because they have some ongoing contact with judges and are arguably more competent to evaluate judges' performance, the polling of police is not as readily understood. Indeed, there was, and continues to be, debate about whether law enforcement officers and probation officers, or indeed any other interest group, should be given the opportunity to participate specially in this kind of process. Interestingly, the Bar Association's evaluation and the police officers' evaluations tend to track fairly closely with each other, lending credibility to and validating both polls in the minds of the public.

We also polled jurors that first year. We no longer poll jurors for a variety of reasons, but essentially our analysis showed that jurors really lacked a frame of reference with which to evaluate judges. Their scores tended to be very high and very uniform. The jurors appeared to lack the ability to discriminate because they evaluated only one judge and so had no basis of comparison. At any rate, we have continued to poll the Bar and police officers before each election every two years for judges standing for retention.

The Bar poll and the law enforcement poll ask questions in

twenty-two areas. Respondents are asked to evaluate a judge's performance on a scale of one to five in four general areas: legal skills, impartiality and fairness, temperament and decorum, and administrative abilities. The law enforcement officers do not rate judges in the legal skills area. In addition, law enforcement officers do not evaluate appellate judges. The law enforcement input therefore is for the nonlegal aspects of judicial performance at the trial level.

When the results of the polls are completed, the Judicial Council analyzes the results and makes a recommendation to the public. The Council then turns those results over to the lieutenant governor, and the results are published in the voters' pamphlets.

In 1976, the first year of the system, the Judicial Council opposed one judge based on the results of the polls. That judge was quite concerned, needless to say, and waged a rather fierce media campaign in opposition to the Council. Alaska requires a fifty percent affirmative vote in order for a judge to stay in office. The opposed judge was retained by an extremely high margin.

In 1978, the Council opposed two judges, and those two judges were also retained in office but by very slight margins. In 1980, the Council opposed two other judges, and again those two judges were retained by slim margins.

In 1982, the Council again opposed two judges. Those two judges were, in fact, rejected on retention by the voters. In 1984, the Judicial Council opposed none of the judges—the Council recommended retention of all judges standing for retention. Two judges who would have been up for retention in 1984, however, who had been recommended against in 1980, declined to stand for retention in 1984.

We believe the evaluation process in Alaska, for retention purposes only, has had a certain amount of success. This conclusion is based on surveys we have conducted of voters' responses. According to a voter survey conducted in 1979, about seventy percent of all Alaska voters indicated they "discriminated" in judge voting, that is, they voted both for and against retention of some judges based upon various types of credible public information available on the incumbent's qualifications, including Judicial Council evaluation materials. At the same time, Alaska voters cast ballots at a record rate. We recently did an analysis of the results of the 1982 election. Seventy-five percent of Alaska's registered voters cast ballots and about eighty-five percent of the voters who voted in the congressional election also voted in the judicial elections. The voting rate therefore was very high when compared to

other jurisdictions, reflecting a relatively high degree of interest in judicial elections.

The information obtained from the survey of voters' responses is statistically analyzed and then is made public. This information is also available to the judges. In addition, we write to a number of attorneys individually who have appeared before each judge and ask those attorneys to provide narrative comments on the judge's performance. We summarize those narrative comments and make the comments available to the judges, which serves as useful feedback for many of the judges.

Confidentiality concerns have arisen in two areas: those who provide us with information want assurance it will not be turned over to others, and frequently judges want assurance that health information will be confidential. As far as the anonymity of such comments is concerned, people who fill out the questionnaires and the polls are assured of confidentiality. If somebody wants to comment or raise a problem or concern about a judge and wants to sign her name, however, it is acceptable so that we can get back to that person and find out some other details. We do maintain the confidentiality of the source. A number of judges have expressed concern about providing health information or submitting to physicals for the purpose of evaluation. Their concerns fall into two areas: either it is an invasion of privacy, or information is not guaranteed to be confidential because the media or somebody could request it and then it would be out of the Council's hands.

We recognize and feel strongly that in order for the Council to get information that is useful, we have to be able to assure confidentiality to all sources. By the same token, we are obligated to protect the rights of privacy and confidentiality of the judges whose performance is being evaluated. Therefore, we have established rules, which very carefully protect the confidentiality of all sources.

In conclusion, one reason our system has been effective is that the Council, as a state agency charged with this responsibility, has developed some credibility over the years because it has withstood the tests of criticism of its statistical techniques and confidentiality and fairness. Second, both the police officers and the Bar evaluate at the same time, and the similarity of their results gives credibility to each other and makes the results more acceptable to the public.

Third, the mass dissemination of the polls' results to the public through the voter pamphlets assures that everyone in the state gets the same sort of information. If the public indeed believes the

results are reliable, they will be inclined to make informed decisions.

Finally, we have not had any competition in recent years from other organizations, from the media, from the Bar, or from private interest groups who decide it is their job to evaluate judges. I cannot guarantee that will last forever.

The Council is looking toward some expansion of its activities, including evaluating the performance of retired judges who serve *pro tem*. This is an area which has fallen through the cracks. The orientation of Alaska's system is more toward accountability than self-improvement *per se*. Therefore, the potential exists to experiment further with performance evaluation.

The Honorable Alan Handler

I had the interesting experience of working with the American Bar Association (ABA) in connection with the formulation of its present guidelines on evaluation of judicial performance. The ABA guidelines, in a way, symbolize the quickened pace that attends the whole subject of the evaluation of judicial performance. It is a rather remarkable document when one considers that it represents, in a sense, a consensus of the organized Bar, a sometimes cantankerous body not given to agreement on subjects as sensitive as this one. It is more remarkable when one considers that it represents not only the thinking and the contributions of lawyers as such, but the thinking and contribution of a considerable number of judges who participated in the effort. The ABA guidelines, in our view, can serve as a very instructive frame of reference for any jurisdiction that is considering whether judicial evaluation should be considered, developed, and implemented on a regular and official basis.

Two major questions should be posed and answered when considering a judicial evaluation program. The first question is whether the jurisdiction should adopt the program at all. Are there problems that are surfacing, that are going unresolved, that are festering, that might be suitably addressed by such a program? Second, how does one go about undertaking the rather monumental reform reflected in an official judicial evaluation effort? This question, in turn, depends very much upon the institutions and the judicial environment of the particular jurisdiction.

Briefly, I will sketch the history of New Jersey's experience and indicate how we focused upon those particular questions and attempted to answer them. The New Jersey court first turned its attention to judicial evaluation in 1979 when it appointed a three-

person committee to study the subject and make recommendations. The decision to study judicial evaluation resulted from the perception that Bar polls and media polls were proliferating. The polls were considered to be unreliable, arbitrary, inconsistent, and, in many instances, downright mischievous in their impact on the judicial groups that were their objects. Another reason prompting the court to take an independent look at the subject was the realization that the organized Bar was taking concerted steps toward developing its own evaluation programs.

A final concern of the court related to the haphazard ways in which judges might succeed or come to grief in connection with reappointment. In some instances, judges, whom many perceived to have been performing well and ably, were victimized by what appeared to be unfair, unjustified, and unreasoned criticism. The judiciary and others were frustrated and unable to counter meaningfully that kind of criticism and felt in some instances that the reappointment process was skewed by unreliable information.

The committee returned its report in 1979. It recommended in principle that there be a program for the evaluation of judicial performance. The linchpin of its recommendation was that the evaluation of judicial performance should be considered an integral component of a judicial improvement system. The report strongly emphasized that judicial evaluation should not be an end unto itself. It further stressed that the program should focus totally on the well-being of the judiciary and the administration of justice. Because that was the purpose of the program, it was to be designed, developed, administered, and controlled by the judiciary. It was, in the view of the committee, to be a program for and by judges. The report also underscored that any program had to accept, as an ongoing and continuing condition, that it be administered and applied in a way that would not threaten or give the appearance of threatening judicial independence and judicial integrity.

The New Jersey Supreme Court accepted that report. It appointed a permanent committee on judicial performance and authorized that committee to design and develop a program for judicial improvement, emphasizing judicial evaluation as an element of that program. The committee in due time submitted a report outlining its recommendations for such a program and secured the authorization of the supreme court to implement it.

The committee decided to introduce the program on an experimental or pilot project basis. The pilot project lasted for approximately one year. Upon the conclusion of that project, the

committee assessed the results to determine whether the program should be instituted on a permanent and regular basis and to determine the configuration of such a program.

The pilot project involved four vicinages or counties. A total of more than seventy judges was involved in the evaluation effort. It was limited to judges sitting on the trial bench, but in all courts: civil, criminal, juvenile, family, and courts handling specialized proceedings. The committee developed questionnaires with the assistance of lawyers and judges. It met with the judges who were participating in the program to elicit their views as well as to acquaint them with the goals of the project. The program relied on lawyers as the principal source of information although it complemented this source with information concerning performance from appellate judges and jurors. It focused on the major definable areas of judicial performance: legal ability, efficiency or productivity, and comportment.

When the project was completed, over 1,000 questionnaires had been sent out to lawyers and returned. The response rate was considered to be satisfactory by our experts. Several hundred questionnaires were sent out to appellate judges and jurors. All the questionnaires were returned anonymously. Confidentiality was carefully and strictly preserved. The results were aggregated on an individual basis and made available to each of the judges who had been the subject of the project, and, in turn, their comments were elicited. As a result of this process, we have reached the following conclusions. These conclusions represent some answers to the questions originally posed by Ms. Brennan.

One overriding conclusion drawn from the pilot project is that a judicial evaluation program can be a positive tool in the administration of the judiciary. The committee further determined that the evaluation subjects—legal ability, efficiency, and comportment—are widely accepted and should continue to be used. The committee was satisfied that reliable information can be captured and presented in a manner that provides judges with perceptions of levels of performance, furnishes insight into the judge's comparative strengths and weaknesses, and serves to identify areas of needed improvement.

The committee concluded that lawyers' evaluations were reliable, responsible, cooperative, fair, knowledgeable, and objective. The committee also concluded that appellate judges were extremely valuable in terms of evaluating trial performance, but their heavy professional commitments must be accommodated in order to elicit their full cooperation. Reflecting the experience of

Alaska, the committee also found that juror responses tended to be of marginal utility because of the "halo effect."

The primary purpose of a judicial evaluation program is judicial self improvement. Much debate exists over the wisdom of combining judicial evaluation with judicial discipline. The ABA guidelines have rejected any correlation between the two areas and quite emphatically state that judicial evaluation should not be used in connection with judicial discipline. This attitude reflects the assumption that a lot of unseemly material might be generated by an evaluation effort, and it would be extremely unfair to a judge if she or he were exposed to discipline. The ABA guidelines suggest that there be no interrelationships between judicial evaluation, which is designed to lead to self-improvement, and judicial discipline.

The ABA guidelines are based on the premise that judicial discipline necessarily involves formal procedures. Discipline is triggered by the filing of some type of complaint or statement which is regarded as a complaint, and that necessarily triggers some kind of formal process. I agree with the ABA's conclusion based on that premise.

In our jurisdiction, however, we have seen that evaluation information can be constructively used in a disciplinary setting. One example is a recent New Jersey experience. A judge was publicly reprimanded for acts of intemperance when handling juvenile matters. On appeal, the judge's attorney asked the court to consider the judge's pilot project evaluations. Unfortunately, we were unable to give evidentiary weight to that information because at that point we had not satisfied our conditions for giving evaluation information sufficient credibility. What is important and instructive is that the data disclosed this judge had been generally regarded as an excellent, conscientious, and scholarly judge who for the most part handled the difficulties of a high-volume, pressure-laden court with considerable skill. The disciplinary committee in that case, however, did not have the opportunity to consider that kind of mitigating information. In my own view, this sort of information is going to be beneficial to an overwhelming number of judges, not only in this type of situation, but also in other ways in the future.

Janet Franz

When addressing any national group, I like to describe Vermont because it is a little different in scope than other states. To the NAWJ in particular, I would like to mention that we just

elected our first woman governor last fall, several months after our first woman trial judge was appointed. I hope this election and appointment are evidence of a trend.

Vermont is a small state with a population of half a million. We are a very rural state. We have only two areas that call themselves cities. Only one, Burlington, can qualify for federal small cities grants, and it has to reach.

Most of our courts are one-judge courts or sometimes quarter- or half-time judge courts. We do not have large urban courts that have several courtrooms running simultaneously, which makes for a little different administration of the system. We are small and still very much a people-to-people state. Therefore, we can sometimes try things on an experimental basis that larger states might not attempt. These characteristics also contribute to our ability to implement changes quickly.

About two and a half years ago, our supreme court appointed a committee to study methods for evaluation and improvement of judicial performance. This committee was formed as a result of a judges' poll conducted by the Young Lawyers' section of the Vermont Bar Association. The chief justice of the Vermont Supreme Court was concerned that the results of the poll were not statistically accurate and, in fact, that this poll was something the judiciary ought to be doing. The chief justice believed that judges should be concerned about and work toward improving their judicial performance. The committee worked for about a year and in August of 1984 made some recommendations to the supreme court, all of which were accepted.

As a result of the committee's effort, a program for evaluation of trial judges was developed and an advisory committee was established to oversee the program. In addition to the trial judge evaluation program which is currently in operation, the committee is studying an evaluation of probate judges and supreme court justices. The evaluation program for the twenty-four trial judges is now ten months old. The committee spent most of its time discussing the aspects of the trial judges' evaluation, and therefore most of its report to the supreme court revolved around those evaluations. The committee focused on trial judges for two reasons. First, those are the most visible judges in the system and the setup in Vermont lends itself to a system of performance review for them that made our job a little easier. Second, in Vermont these judges are appointed by the governor. The committee, therefore, did not worry that if we made a mistake, such as judging

something wrong in our data collection, we were going to be influencing an electorate.

The committee that designed the evaluation program consisted of a supreme court justice, a probate judge, two trial court judges, and the administrative judge responsible for supervising the trial judges. Other committee members included a member of the Vermont Legislature's Joint Committee on Retention, a court clerk, and an attorney, who coincidentally was appointed to one of the ABA task forces and worked with Ms. Brennan. As a result, a very nice dialogue occurred between the state and ABA groups which was helpful to us in Vermont.

Very early on the committee, and later the supreme court, agreed that the purpose of any evaluation program in Vermont should be improved judicial performance. The purpose of this evaluation program was not for retention decisions, not for appointment decisions, and not for public relations (which could, in my mind, be a legitimate use). The most important and primary goal would be improved judicial performance, and most of our other work flowed from that decision.

When designing a judicial evaluation instrument, the first and most important question is, what criteria should be used to decide whether a judge is doing a good job or not? We hired a management consultant to help us answer that question. I believe Connecticut did something similar. Connecticut, in fact, had some corporation people on its committee to discuss with them private sector evaluations. In trying to answer the question, we listed what we felt a judge was expected to do and from there asked the question, who would have information in these various areas as to a judge's actual performance?

Next we tried to devise some very specific questions so that the responses would tell us whether a judge was or was not doing what the judge was expected to do. We later grouped those questions into areas of responsibility. We narrowed the areas of responsibility, concentrating on administration, legal skills and preparation, conduct of court proceedings, and communication.

We wanted to collect objective data. One of the ways that we tried to keep data free from personal bias, as well as sex or racial bias, was by asking for very specific information. We did not ask, how would you rate this judge's demeanor—excellent, average, or poor? Instead, we tried to ask very specific questions such as, when this judge is presiding, does court begin on time? It is difficult to inject bias into that question. While being specific does not solve the problem of bias in the responses, specific questions at

least mitigate bias. It is much harder, however, to filter out bias in questions on legal skills and knowledge.

As in most states, we asked questions primarily of attorneys because they are before judges more than anyone else. We also polled jurors and found, as have other states, jurors almost invariably liked their judge regardless of who that judge was. Although jury polls have been of limited value, we are continuing to do jury polls for two reasons: partly because the program is only ten months old, and partly because the results of jury polls could point out any red herrings, particularly bias, and give extra weight to attorneys' comments. For example, if attorneys say a judge has a strong bias toward the defense, and jurors say the same thing, a very strong case of bias exists. In addition, only the jurors may notice any bias toward or against attorneys. We are beginning to experiment with questionnaires to litigants, but it is too early to comment on this aspect of the judicial evaluation program.

At one point we talked about polling court staff. Since our trial judges rotate, we could poll court staff more easily than some states and still preserve anonymity. The results of court staff polls could be kept anonymous because a judge in the course of the year sits in two, three, or more courts and each court has a court reporter, court officers, as well as clerical staff. We have not yet polled court staff on a full scale basis but instead use court staff as a source of informal feedback on judges. All of the information collected from attorneys and jurors goes to the administrative judge who oversees the work of the trial judges. If he sees any areas that seem questionable—"Boy, I've never heard this before, this really surprises me" or "I'd like to know more about this," he then may ask the clerk if she or he knows anything more about this area or if anyone else on the staff might have information. Thus, the administrative judge is using the court staff more informally. The evaluation program allows the administrative judge to go to any source for information, either formally or informally.

The administrative judge system in Vermont is advantageous for implementing this type of program. Since we wanted to address judicial performance improvement, it was important that the trial judge get feedback on performance. We also wanted the evaluation to be conducted, however, in the most confidential manner. The system that we designed allows both feedback and confidentiality. The results of the polls go to the administrative judge, who makes an appointment with the trial judge, and they sit down to discuss the results. They are the only two people in the system

who have access to those results, and as soon as their meeting is over the results are destroyed.

So far, five judges have completed the evaluation process. We instituted a rotating system. With only twenty-four trial judges, we decided to establish a quarterly system, where three judges are evaluated every three months. In any three-month period, the highest number of questionnaires that an individual attorney could receive would be three. We felt that we would get quicker responses and better quality of responses from attorneys who were asked to evaluate three judges than if attorneys were asked to evaluate all judges in the state at one time.

Another key piece of the evaluation program is a self-evaluation questionnaire that the trial judge fills out before her meeting with the administrative judge. It is one of the most important elements of the evaluation because it includes both a self-assessment component as well as a goal-setting component. Questions include: How would you describe your settlement techniques? Are you fairly strong at settlement techniques? Do you wish to change the settlement techniques you use? This self-evaluation questionnaire has significant implications for judicial programs. By discussing their present capabilities as well as future goals with the administrative judge, the trial judges are able to assess how their goals can be practically reached. For example, some judges may need to take extra courses while others may need additional law clerk time.

The goal of the evaluation system is to collect information from the judge and from persons familiar with the judge and to use that information in a way that facilitates improved judicial performance. So far, the judicial evaluation program is working. We receive many written comments in addition to the information requested on the questionnaires. I have not heard even so much as an informal conversation that would indicate to me there have been any breaches of confidentiality. Nor, to touch on one of Ms. Brennan's other questions, have I heard any judges say anything about the evaluation hampering their discretion or independence.

We discussed at length whether or not this information should be available to the Joint Committee on Retention of the Vermont Legislature, which makes recommendations to the whole legislature every six years. Our strong belief was that as noble an objective as that would be, that was not the purpose of the program. We did not think both purposes could be served. We were afraid that if the results were given to the retention committee the results would then have to be given to the entire Legislature and

would then be public domain. Attorneys would not be candid if they knew that whatever they said could be used against (or for) them by a judge.

The Joint Committee on Retention wrote to us and asked whether information obtained from the judicial performance evaluations would be available to the committee. We wrote back and said no, stating that aiding retention determinations is a very good purpose for information of this sort, but our information would not be available due to concerns over confidentiality. I do not know whether it is the result of our program or our correspondence, but the Vermont Legislature passed a bill in 1985 under which the retention committee is starting its own program of polling attorneys to get information for retention decisions. The retention committee is in touch with the judicial performance advisory committee in order to obtain advice on problems they encounter. I am willing to support the retention committee's efforts to conduct their own evaluation, because this enables our system to continue for the purpose it was designed. Both systems could turn into a very comprehensive judicial evaluation system.