

APPENDIX I
AMERICAN BAR ASSOCIATION
GUIDELINES FOR THE
EVALUATION OF JUDICIAL PERFORMANCE
WITH COMMENTARY

FEBRUARY 2005

I. GOALS

Guideline 1-1. Judicial evaluation programs improve the performance of individual judges and the judiciary as a whole. All court systems should develop and implement a formal program for the evaluation of judicial performance.

Commentary

The improvement of judicial performance is a goal shared by all. Programs for evaluating the performance of sitting judges provide a basis for judges to maximize their potential for excellence through self-improvement without jeopardizing judicial integrity and independence. Evaluation programs give judges the feedback they need to improve their own performance and that of the courts on which they serve.

Guideline 1-2. In jurisdictions where judges are subject to reappointment, retention, or reelection, judicial evaluation programs enable those responsible for continuing judges in office to make informed decisions.

Commentary

Programs for evaluating judicial performance provide an assessment of the individual judge's competence in office. In all but three states, judges are subject to reappointment, retention, or reelection. Evaluation programs should supply the reliable and unbiased information about judicial performance that is needed to make sound judgments regarding the continuation of judges in office.

II. USES

Guideline 2-1. Primary uses of judicial performance evaluation include promoting judicial self-improvement, enhancing the quality of the judiciary as a whole, and providing relevant information to those responsible for continuing judges in office.

Commentary

These guidelines are not intended to be used where applicable law or regulation excludes judges from performance appraisals.

Guideline 2-2. Additional uses that may be considered include the effective assignment of judges within the judiciary and the improved design of continuing education programs.

Commentary

In some jurisdictions, judges with administrative responsibilities and/or court administrators are charged with assigning judges to either the trial or appellate level, civil or criminal cases, or courts with general or specialized jurisdiction. The information obtained through judicial evaluation programs will aid those responsible for making such assignments. Evaluations of judicial performance will also allow those who design continuing education programs to identify and target resources to areas where increased education would be most beneficial.

Guideline 2-3. The uses of judicial performance evaluation do not include judicial discipline. The information developed in a judicial evaluation program should not be disseminated to authorities charged with disciplinary responsibility, unless required by law or by rules of professional conduct.

Commentary

With respect to the relationship between judicial evaluation programs and judicial discipline, lawyers and judges who participate in judicial performance evaluations should be aware of and act according to their obligations under their states' rules of professional conduct and codes of judicial conduct. Whether or not information relevant to judicial discipline ought to be turned over to judicial disciplinary bodies or be privileged and not subject to subpoena is an issue that each jurisdiction must resolve for itself.

III. DISSEMINATION

Guideline 3-1. The dissemination of data and results from a judicial evaluation program should be consistent with and conform to the uses of the program. Except for the authorized uses of the performance evaluation and consistent with the law, the data and results should be confidential.

Commentary

In developing judicial evaluation programs, jurisdictions should be aware of applicable public disclosure laws and take steps to resolve any conflicts between maintaining the confidentiality of evaluation results and public disclosure requirements.

Guideline 3-2. When judicial evaluations are used only for judicial self-improvement, individual results should be provided only to the judge evaluated and the presiding or supervisory judge responsible for the performance of the court on which the judge serves.

Commentary

Consistent with the goal of improving judicial performance, the individual judge and the presiding or supervisory judge should receive complete results and supporting data concerning the individual judge. In order for effective use to be made of this

information, the individual judge and the presiding or supervisory judge also need a frame of reference against which to assess the judge's levels of achievement. It is therefore important that judicial performance data be tabulated in some aggregate form for all of the judges participating in the particular evaluation effort.

In many jurisdictions, respondents to performance evaluation surveys and members of the general public are encouraged to provide narrative comments pertaining to a judge's performance. When narrative comments are solicited during the evaluation process, such comments should be made available only to the evaluated judge and the presiding or supervisory judge.

Guideline 3-3. When judicial evaluations are used to improve the quality of the judiciary as a whole, results should not identify or give comparative rankings of individual judges.

Commentary

In order to improve the overall quality of the judiciary, evaluation information may also be provided to the highest court or other body responsible for judicial administration. Evaluation results used for this purpose should consist of summary data for the entire court without identification of individual judges. Similar information should be provided to those responsible for designing continuing judicial education programs.

Comparative rankings of individual judges are inappropriate. However, in some instances comparisons among judges may be helpful. In order to demonstrate to a judge the need for improvement within a certain category, it may be necessary to make some comparison to the evaluations that other judges received in the same category.

Guideline 3-4. When judicial evaluations are used to inform decision makers regarding the continuation of judges in office, results should be made readily available to those responsible for continuation decisions, including voters, governors, legislatures, and commissions.

-4.1. Those responsible for reappointing, reelecting, or retaining judges should be provided with objective summaries of evaluation results for each judge and an explanation of how to interpret the results.

-4.2. If evaluation results are provided to an individual or entity responsible for continuation decisions, and those results include assessments of a judge's overall performance or recommendations as to whether a judge should be continued in office, judges should have an opportunity to review and respond to the evaluation report before it is disseminated.

-4.3. If evaluation results are publicly disseminated, and those results include assessments of a judge's overall performance or recommendations as to whether a judge should be continued in office, judges should have an opportunity to review, respond, and meet with members of the evaluation body before the results are made public.

Commentary

In all but three states, judges are subject to reappointment, retention, or reelection. In some jurisdictions, the electorate, the governor, or other entity responsible for deciding whether to retain a judge in office may have only limited or unreliable information on which to base decisions. Objective summaries of evaluation results for individual judges, along with an explanation of how to interpret the results, should be made available to those responsible for continuing judges in office. Results should not be disseminated in a format that would readily allow the public or the media to compare or rank individual judges.

In most jurisdictions that use judicial performance evaluations to inform decision makers regarding the continuation of judges in office, a recommendation is made as to whether judges should be continued in office or an assessment is provided as to whether judges meet or do not meet performance standards. In these jurisdictions, judges should be allowed to look over their evaluation reports before they are disseminated publicly and to meet with the evaluation committee or a representative subcommittee (i.e., composed of a judge, lawyer, and non-lawyer) in order to discuss the reports.

When the continuation decision is made by the electorate, ensuring that voters receive the necessary information will be more difficult than informing governors, legislatures, or commissions. In states where judges must be reelected or retained by the voters, performance evaluation results should be disseminated as widely as possible, including by mail, on the Internet, in state and local newspapers, and/or in public facilities such as libraries and courthouses.

IV. ADMINISTRATION AND SUPPORT

Guideline 4-1. Ultimate authority over the development and implementation of a judicial performance evaluation program should be vested in the highest court or other constitutionally mandated body having ultimate responsibility for judicial administration.

-1.1. In states where performance evaluation programs have not been established by the judiciary or other governmental body, bar associations should develop and administer evaluation programs according to these guidelines.

-1.2. In states where judges are chosen in contested elections, it may be inappropriate for the judicial branch or any other entity using public funds to disseminate performance evaluations of incumbent judges running for reelection. In order to provide voters in these states with relevant information, bar associations should develop and administer judicial performance evaluation programs according to these guidelines.

Commentary

The judiciary should take the institutional responsibility for developing and overseeing a judicial performance evaluation program. The court or body with supervisory authority over the judiciary is best suited to carry out this function. In some states that evaluate judges, an agency called the judicial council is responsible for designing and

implementing the evaluation program. Judicial councils are created in these states' constitutions and are given authority over various aspects of court administration.

An exception to this principle may exist in states where judges run in contested elections. Concerns may be raised by the fact that a government entity is using public funds to evaluate and inform voters regarding one candidate – the incumbent judge – but not the other – the challenger. In contested election states, the judiciary should develop and implement evaluation programs for judicial self-improvement and the improvement of the judiciary as a whole, but bar associations in these states are better positioned to evaluate judges for the purpose of informing the electorate. Bar associations should also conduct evaluation programs consistent with these guidelines in states where an official evaluation program has not been established.

Guideline 4-2. The day-to-day activities of the judicial evaluation program should operate through an independent, broadly based, and diverse committee.

-2.1. In jurisdictions where judicial evaluations are used solely for self-improvement and for improving the quality of the judiciary as a whole, oversight committees should be composed of members of the bench and the bar.

-2.2. In jurisdictions where evaluations are used to inform decisions regarding the continuation of judges in office, oversight committees should also include members of the public who are familiar with the judicial system.

Commentary

The continuing administration of the judicial evaluation program should be the responsibility of an independent committee specially created by the entity having ultimate authority over the program. At a minimum, evaluation committees should include both judges and lawyers, as these groups are most familiar with the work and responsibilities of judges. Since the improvement of judicial performance is a matter of public interest and importance, representatives of the public may also serve on evaluation committees. Committee members should be drawn from throughout the relevant jurisdiction and should reflect the diversity of the jurisdiction's population.

Guideline 4-3. Staff support and adequate funding should be available to support a judicial evaluation program of high quality.

Commentary

In order to be successful, a quality program for the evaluation of judicial performance must be adequately funded. In part, the level of funding necessary may depend on the skills of existing staff and the availability of other resources. Survey research experts should be utilized in the drafting of questionnaires and the tabulation and analysis of results. If such experts are not already on staff, funds should be available to contract for their services. Staff members, such as a court administrator's or clerk's office, may handle the day-to-day work of distributing and collecting questionnaires.

Guideline 4-4. Judicial evaluation programs should be structured and implemented so as not to impair judicial independence. The evaluation process should be free from political, ideological, and issue-oriented considerations.

Commentary

The preservation of judicial independence must be recognized in the development and administration of any judicial evaluation effort. From the wording of the questionnaire to the identification of respondents to the dissemination of results, performance evaluation programs should be conducted so that judicial evaluations are not based on agreement or disagreement with a particular judicial philosophy or case outcome.

Judicial evaluations based on appropriate criteria and reliable and valid methodology, as described in Sections V and VI of these guidelines, pose no threat to the independence of the judges being evaluated.

Guideline 4-5. Judicial evaluation programs should be developed systematically and may be implemented in progressive stages. Evaluation programs should remain flexible so that they may be modified as needed. The entity having ultimate responsibility for the evaluation program should conduct periodic assessments of the program.

Commentary

Judicial performance evaluation programs require careful consideration and systematic planning for successful implementation. Periodic review of an ongoing program is important and highly desirable. As such programs are implemented, expertise based on experience will develop. Debate, experimentation, and the publication of research will suggest new ways in which to pursue the objective of improving judicial performance. This suggests that flexibility be maintained in the program design so that modifications can be made where necessary.

Experience indicates that it may be advisable to implement the program in progressive stages. A given jurisdiction may outline a complete program of judicial evaluation and then specify that certain phases be implemented in the first year, additions made in the second year, and so forth. Judges should be involved in each step of the development, implementation, and oversight of the judicial evaluation program. Such an approach not only capitalizes on the knowledge and expertise of judges, but also ensures that evaluation programs will not unduly interfere with the regular performance of judges' duties or infringe upon judicial independence.

V. CRITERIA

Guideline 5-1. A judge should be evaluated on his or her legal ability, including the following criteria:

- 1-1. Legal reasoning ability.**
- 1.2. Knowledge of substantive law.**

-1.3. Knowledge of rules of procedure and evidence.

-1.4. Keeping current on developments in law, procedure, and evidence.

Commentary

Judges have a duty to know and understand the law. Judges must address, and therefore be knowledgeable in, many areas of law including a broad range of substantive law issues, constitutional law, procedure, and evidence. Such knowledge requires that a judge devote time to the study of recent legal developments, attend judicial education programs, and continue the process of a judge's self-education.

Judges need constructive comment with respect to how others view their understanding and knowledge of the law. A judicial evaluation program will assist the judge in assessing his or her knowledge of the law and, where appropriate, permit the judge to devote further time to self-improvement with respect to knowledge of the law.

Where these criteria are used, evaluators must be cautioned to disregard their personal feelings about a judge's decisions. These criteria measure knowledge of the law; they do not measure the extent to which the evaluator and the judge share similar legal philosophies or ideologies.

Guideline 5-2. A judge should be evaluated on his or her integrity and impartiality, including the following criteria:

-2.1. Avoidance of impropriety and the appearance of impropriety.

-2.2. Treating all people with dignity and respect.

-2.3. Absence of favor or disfavor toward anyone, including but not limited to favor or disfavor based upon race, sex, religion, national origin, disability, age, sexual orientation, or socioeconomic status.

-2.4. Acting fairly by giving people individual consideration.

-2.5. Consideration of both sides of an argument before rendering a decision.

-2.6. Basing decisions on the law and the facts without regard to the identity of the parties or counsel, and with an open mind in considering all issues.

-2.7. Ability to make difficult or unpopular decisions.

Commentary

It is essential in any program for the evaluation of judicial performance that integrity be included as a criterion. Avoiding the appearance of impropriety is as important as possessing integrity itself. In deciding cases, judges should attempt to avoid all biases that could impair, or be seen as impairing, their judgment.

Judges must not show favor or disfavor with respect to issues, parties, or attorneys in matters before the court. Judicial impartiality is defined by the 1990 ABA Model Code of Judicial Conduct (Revised 2003) as an absence of bias or prejudice in favor of, or against, particular parties or classes of parties, and an open mind in considering issues that may come before the judge. Judges must appear to be and actually be fair. Both the appearance of and the quality of fairness are essential. For example, the fairness of a judge who consistently disregards or overrules one side's objections may be questioned, even if the rulings are legally sound.

Judges must be aware of perceived and actual biases so that they can avoid any such tendencies. Judicial evaluation programs can assist judges in identifying biases based on sex, race, ethnicity, economic status, and other personal characteristics and preferences.

Good judges have the ability to decide cases on the basis of the applicable law and facts without favor or disfavor based on the identity of the parties or their counsel. Good judges are also willing to rule on issues without regard for the popularity of their rulings and without concern for or fear of criticism.

Guideline 5-3. A judge should be evaluated on his or her communication skills, including the following criteria:

-3.1. Clear and logical oral communication while in court.

-3.2. Clear and logical written decisions.

Commentary

All judges must be able to communicate effectively. Effective communication skills include the ability to speak and write so that what is expressed is understood. The law of a case, whether presented in spoken or written form, should be clear and concise. If it is not, the law remains ambiguous not only for the parties to the case, but also for other litigants who attempt to employ it as precedent.

Judges must also recognize the potential negative impact of verbal and nonverbal communications such as tone of voice, facial expressions, eye contact, hand motions, and posture. These mannerisms can create either an appropriate or inappropriate atmosphere in a judicial proceeding. Avoiding negative verbal and nonverbal communications is equally important.

Guideline 5-4. A judge should be evaluated on his or her professionalism and temperament, including the following criteria:

-4.1. Acting in a dignified manner.

-4.2. Treating people with courtesy.

-4.3. Acting with patience and self-control.

-4.4. Dealing with pro se litigants and litigation fairly and effectively.

-4.5. Participating and providing leadership to an appropriate degree in professional development activities and in jurisdiction-wide and statewide court improvement and judicial education activities.

-4.6. Promoting public understanding of and confidence in the courts.

Commentary

The image of a judge is important, for a positive image creates respect for the judiciary. The most important element of a positive image is that elusive quality called judicial temperament. It includes such criteria as patience, courtesy, dignity, and compassion.

Patience and courtesy are essential qualities for all judges, but particularly for a trial judge. Striking an appropriate balance between patience and courtesy, and a trial judge's authority in presiding over a case, is not always an easy task. For instance, while a judge should exercise control in the courtroom so as to expedite the proceedings, such "control" should not interfere with a lawyer's right to present a case or a pro se litigant's right to represent him or herself.

While appropriate courtroom demeanor is possibly more important for a trial judge, in view of a trial judge's frequent interaction with the public, it is also an important criterion for an appellate court judge. This criterion asks the evaluator to measure how well the judge listens and whether the judge is fair and courteous to counsel. At the appellate level, this criterion is, for the most part, limited to oral argument.

Judges should also seek opportunities for and participate in professional development programs. Professional development opportunities can take a variety of forms. Continuing legal education programs contain many topics that are best taught through the views and experience of judicial office. Similarly, law schools need judges to present a practical perspective to students, particularly in clinical practice programs. There are also a number of state and national programs for continuing judicial education, and many states require judges to complete a minimum number of such courses.

The improvement and effective administration of the courts require judges to participate and provide leadership to an appropriate degree at both the local and statewide levels. The effective management and operation of local courts cannot be left entirely in the hands of a presiding or supervisory judge or of staff; it depends upon appropriate involvement of all judges within the jurisdiction. Most states have statewide judicial conferences or institutes responsible for providing technical assistance, substantive materials, and education programs for judges. The effectiveness and success of these and similar activities depend upon the active involvement of judges.

Many judges give of their time and efforts to improve the judicial system in other ways. These include public speaking, community education programs, working with committees and other groups to further the quality and/or understanding of the judiciary, and other related activities. Judges should also be encouraged to speak on

matters directly affecting the administration of justice, while being careful not to prejudice themselves on substantive legal issues that may come before them.

Guideline 5-5. A judge should be evaluated on his or her administrative capacity, including the following criteria:

- 5.1. Punctuality and preparation for court.**
- 5.2. Maintaining control over the courtroom.**
- 5.3. Appropriate enforcement of court rules, orders, and deadlines.**
- 5.4. Making decisions and rulings in a prompt, timely manner.**
- 5.5. Managing his or her calendar efficiently.**
- 5.6. Using settlement conferences and alternative dispute resolution mechanisms as appropriate.**
- 5.7. Demonstrating appropriate innovation in using technology to improve the administration of justice.**
- 5.8. Fostering a productive work environment with other judges and court staff.**
- 5.9. Utilizing recruitment, hiring, and promotion policies and practices to ensure that the pool of qualified applicants for court employment is broad and diverse.**
- 5.10. Acting to ensure that disabilities and linguistic and cultural differences do not limit access to the justice system.**

Commentary

Although managerial responsibilities will vary among jurisdictions and assignments, judges at any level and with any assignment must possess good management skills in order to be productive. A judge cannot be effective without the ability to organize, manage, and effectively control judicial proceedings.

Productivity is a function of time management – that is, how well the judge uses his or her time. A judicial evaluation program should help a judge assess his or her ability to settle cases, to be prompt in issuing decisions, and to function both efficiently and effectively.

Promptness in adjudication is essential and is equally important at both the appellate and trial levels. While an appellate judge may be unable to control completely the promptness with which a case is decided and an opinion filed, a reasonably prompt decision in all cases is important, and every appellate court judge must work toward this end. Trial court judges are generally in a position to control the disposition of their cases without regard to the schedule of other judges.

A jurisdiction developing a judicial evaluation program for self-improvement can help determine specific time limits for completing assignments, where appropriate (e.g., 90 days, 120 days). At a minimum, however, promptness includes starting judicial proceedings on time and ending them on time.

One aspect of efficient case management is promoting the use of settlement and other alternative dispute resolution mechanisms to resolve legal conflicts. At appropriate stages during pre-trial proceedings and before resolution of a case, and within ethical boundaries, judges should encourage the parties to reach a settlement. Judges should also make the parties aware of such options as mediation, arbitration, and collaborative law.

The use of technology in the courtroom can enhance services and reduce costs for courts, lawyers, and litigants. The innovative use of technology, with appropriate security controls, can improve the administration of justice. Examples include providing electronic access to court records, opinions, and calendars; allowing electronic filing of court documents; and enabling teleconferencing and videoconferencing for pre-trial proceedings. Judges have a responsibility to inform themselves about available technologies, to incorporate their use as appropriate, and to keep current on new developments.

Judges must possess the attribute of cooperativeness. Trial court judges frequently work with other trial court judges and share responsibility for administration, scheduling, and other tasks crucial to accomplishing the work of the trial court. Appellate judges, who almost always work on panels with other judges, must freely and effectively exchange views with other judges and negotiate differences that arise. Judges at both levels of court also interact regularly with court personnel. Promoting cooperative relationships with other judges and with court staff will create a harmonious and productive work environment.

In order to preserve the public's trust and confidence in the judiciary, judges should be vigilant in ensuring that disabilities and linguistic and cultural differences do not limit access to the courts. They should also take advantage of the steps available to them to promote meaningful diversification of the judicial branch.

Guideline 5-6. Additional criteria should be developed reflective of jurisdiction (specialized versus general) and level of court (trial versus appellate).

-6.1. A specialized court judge should be evaluated according to whether he or she demonstrates the knowledge and skills necessary.

-6.2. An appellate court judge should be evaluated on the quality of his or her preparation for and participation in oral argument and on his or her effectiveness in working with other judges of the court.

Commentary

Many judicial systems include specialized courts such as those handling matters pertaining to small claims, probate, juvenile, tax, family relations, or administrative law. These courts have varying functions, depending on the particular jurisdiction. Each jurisdiction planning for judicial performance evaluation should give careful consideration to its specialized judges to determine whether it is appropriate to develop additional criteria that would be reflective of the unique problems and requirements of specialized judges.

Developing additional evaluative criteria may also be appropriate for appellate court judges. Judges who serve on appellate courts work with other judges and generally make decisions collectively rather than individually. In this environment, it is essential for judges to be able to resolve differences with their colleagues, to consider the views of others, and to offer constructive criticism with respect to each other's draft opinions. While only other appellate judges may be able to evaluate this behavior, these are important criteria, particularly in a program for self-improvement.

VI. METHODOLOGY

Guideline 6-1. The judicial evaluation process is comprised of data collection, synthesis and analysis, and its usage.

Commentary

The evaluation process, once its design has been completed, can be divided into three parts. First is data collection. This involves gathering the raw data, from whatever sources, on which the evaluation is to be based. Such data may be collected on a continuing or periodic basis. The next step in the process is the synthesis and analysis of the data. This involves organization of the raw data in a useful fashion so that it can be readily understood and put to the use for which it is intended. Finally, the process includes the dissemination and use of the data. These aspects of the evaluation process are addressed in Sections II and III of these guidelines.

Guideline 6-2. Expert competence should be used in developing methods for evaluating judges and collecting and analyzing data.

Commentary

It is strongly recommended that experts be used in designing approaches, instruments, and techniques to be used in judicial evaluation programs. Experts have distinctive qualifications in terms of identifying information sources, drafting questionnaires, and assessing the reliability of evaluation methods. It is important, however, that those most familiar with judicial performance—judges and lawyers—work closely with experts to maximize the experts' understanding of the subject matter and the goals of the program.

Most judicial evaluation programs solicit performance information through questionnaires completed by individuals who have had professional contact with

judges. Proper construction of these questionnaires is critical to obtaining unbiased evaluations, and experts should be utilized to facilitate this task.

Program administrators should pilot test the evaluation instrument prior to its broad use for evaluative purposes. Pilot testing allows assessment of the reliability and validity of the instrument, testing of procedures, and observation and training of program staff. Comments and suggestions should be solicited on both the instrumentation and the evaluation process itself in order to refine and improve upon the original design. Expert assistance is particularly important at this stage of the process.

Guideline 6-3. Behavior-based instruments should be used to evaluate judges.

Commentary

The evaluative portion of questionnaires should ask respondents to assess a judge's performance with respect to actual behavioral examples instead of general qualities such as legal ability or temperament. For example, rather than asking respondents to evaluate a judge's impartiality, respondents may be asked whether a judge "treats individual parties with dignity and respect notwithstanding their appearance, lifestyle, or personal views," or "reveals personal bias toward one side in a case." These types of questionnaire items can generate more meaningful information about judicial behavior.

Behavior-based instruments avoid some of the biases that commonly affect survey questionnaires. There are a variety of sources from which bias may arise in judicial evaluation questionnaires. Scaled items are used in many jurisdictions, with respondents indicating degrees of agreement or satisfaction with statements about a judge's behavior. For example, respondents may be asked to characterize a judge's "sense of basic fairness and justice" as unacceptable, deficient, acceptable, good, or excellent; determine whether a judge "was punctual in convening court" consistently, occasionally, or never; or assign a judge's "courtesy toward pro se parties" a grade of A, B, C, D, or Fail. A potential source of bias in such techniques comes from the fact that each respondent must define the available responses for him or herself. With varying standards regarding what behaviors are unacceptable, occur with occasional frequency, or are deserving of a B grade, two respondents may observe the same behavior and evaluate it differently.

Other sources of bias include a propensity for respondents to evaluate all judges in the middle, known as the central tendency bias; to give every judge high marks or to evaluate each one too harshly, known respectively as the leniency error or the stringency error; to allow their assessment of a judge on the first dimension to determine their assessments on all other dimensions, known as either the halo effect or the horns effect according to whether the initial assessment is positive or negative; and to avoid giving responses at either end of a scale, known as scale shrinking. Techniques exist to identify some of these biases in completed questionnaires. When such biases are found to exist, thoughtful judgments may then be made regarding the handling of those questionnaires.

A different type of bias may be found in the disparate performance evaluations received by men and women judges. According to a 1995 report of the ABA's Commission on Women in the Profession, "[j]udicial evaluation programs reflect that women judges endure consistently stronger criticism than their male colleagues, especially in subjective categories such as demeanor." For example, a 1993 analysis conducted by the Colorado Judicial Performance Evaluation Commission found clear gender-related bias in lawyer evaluations of judges. Similar conclusions have been reached by state task forces on gender bias in the courts. An additional benefit of behavior-based evaluation instruments is that questionnaire items reduce subjectivity in assessments of judicial performance, thus limiting the potential for gender and other biases to influence responses. In a behavior-based evaluation instrument, items relate to judges' actual behaviors rather than characterizations of judges' actions as proper or improper.

Guideline 6-4. The evaluation process must ensure the anonymity of individual respondents.

Commentary

Regardless of the intended use of the performance evaluation, the program must ensure the anonymity of respondents to performance questionnaires and of members of the public who submit narrative comments. The reliability of the evaluation process depends on the willingness of evaluators to provide candid and honest feedback without fear of reprisal.

In some jurisdictions, the anonymity of individual respondents is ensured through the same method used by researchers who conduct anonymous mail surveys. Along with the performance questionnaire, respondents are sent an envelope in which to seal the questionnaire for its return. This envelope is placed inside another envelope that is marked with the respondent's name and an identification number and mailed to the evaluation committee. The information on the outer envelope is used to verify that the respondent has returned the questionnaire. The outer envelope is then destroyed before the inner envelope is opened. Methods such as this not only ensure the anonymity of respondents, but also make it possible for program administrators to track and follow up with those who have not yet responded.

Protecting the anonymity of respondents is of particular concern in smaller jurisdictions and in jurisdictions where prosecutors and/or public defenders are assigned to particular courts. Program administrators in these jurisdictions should consider extending data collection periods, limiting the demographic information that is requested of respondents, and/or providing limited or aggregated results to prevent identification of respondents.

Guideline 6-5. Reliable sources of information should be developed for judicial evaluation programs.

-5.1. Multiple sources should be used whenever feasible.

-1.1. Potential sources of information for trial judge evaluations include attorneys, jurors, litigants, and witnesses who have appeared before the judge; non-judicial

court staff, social service personnel, and law enforcement officials who have had regular contact with the judge; and appellate judges who have reviewed the judge's decisions.

-1.2. Potential sources of information for appellate judge evaluations include attorneys who have appeared before the judge, non-judicial court staff who have had regular contact with the judge, other appellate judges, and trial court judges whose decisions have been reviewed by the judge.

-5.2. Sources should be limited to those with personal and current knowledge of the judge.

-5.3. Objective sources of information may include public records.

Commentary

The effectiveness of any judicial evaluation program will depend in large part on the reliability of the information it generates. One of the primary duties of an evaluation program is to identify reliable sources of information.

Whenever feasible, multiple sources of information should be used. To the extent that these sources represent persons with different perspectives, they may each provide information that is not available from the other. To the extent that the sources have the same basis for evaluation, results from the different sources can be compared and used to validate the results. To enable judges to assess their performance among various demographic groups, demographic information including gender, age, and race/ethnicity should be solicited from those who evaluate judicial performance. The value of examining a judge's evaluation results across various demographic groups will be limited where a sufficient number of respondents have not provided this information.

Traditionally, the chief source of evaluative information concerning judges has been lawyers. Lawyers are particularly well qualified to evaluate judges. However, for trial judges, most jurisdictions go beyond surveying attorneys and question other individuals who come into contact with judges. Jurors, witnesses, litigants, and court personnel have firsthand knowledge of judicial conduct that is highly relevant to performance evaluation. Social service personnel and law enforcement officials may also have regular contact with trial judges and should be included in the evaluation process.

Peer evaluations may also be beneficial. Although trial judges rarely observe the courtroom work of their colleagues, appellate judges are in an ideal position to provide evaluative information about trial judges that may not be available from other sources. Appellate judges may also evaluate the performance of their fellow judges. However, due to confidentiality considerations, peer review may not be appropriate for all appellate courts.

Every effort should be made to enhance the reliability of evaluative data by ensuring that respondents have direct and personal knowledge of the judge. In some jurisdictions, respondents are given questionnaires and asked to complete them before leaving the

courthouse. This method is likely to provide higher response rates than would a mail survey. However, this approach may not be appropriate for obtaining feedback from attorneys. Some attorneys appear before a particular judge several times during an evaluation period, making it difficult to ensure that each attorney evaluates a judge only once. Attorneys may also be reluctant to complete questionnaires at the courthouse because of confidentiality concerns. For these reasons, mail surveys may be more appropriate for attorneys.

To guarantee that evaluations are submitted only by attorneys who have appeared before the judge being evaluated, it may not be sufficient to mail questionnaires to all members of the state bar, asking them to opt out if they have not appeared before the judge during the evaluation period, or to the attorneys-of-record in cases over which the judge presided. In some jurisdictions, steps are taken to identify those attorneys who have personally appeared before the judge during the period in question and to mail questionnaires to those attorneys only.

Regardless of the method that is used to distribute and collect performance questionnaires, respondents should be asked to identify the type and extent of contact they have had with the judge being evaluated.

Information should be obtained from respondents while it is relatively fresh. While some may question whether evaluative data is best gathered in close proximity to the event observed or participated in, or at a later date when one may be able to be more dispassionate, there is a point after which reliable information cannot be obtained. As a general rule, evaluations should be based on contacts with judges that have occurred within the past year.

Evaluation programs should include both subjective and objective measures of judicial performance. Public records may provide an objective source of information. Among the objective indicators that many jurisdictions use are case management statistics, peremptory challenges to remove judges, formal disciplinary sanctions against judges, and judges' participation in continuing legal education programs.

One caveat is appropriate with respect to the use of case management statistics. A focus on tangible results such as the number of hearings held, the number of motions or cases decided, or the number of jury trials may be appropriate indicators of a judge's administrative capacity. However, a weighting of factors such as the complexity of a judge's cases and the scope of a judge's jurisdiction should be considered.

Objective components should not provide the sole basis for a judge's performance evaluation; rather, they should be used in conjunction with a multi-faceted evaluation program.

Guideline 6-6. At the outset of the evaluation program, program administrators should establish minimum thresholds for both response rates and number of respondents.

Commentary

In order to ensure the reliability of evaluation results, program administrators should establish minimum thresholds for both response rates and number of respondents. Even when these thresholds are met, evaluation results should include the overall response rate and the number of respondents on which the evaluations are based.

Every effort should be made to obtain as high a response rate as possible. In order to maximize response rates and minimize non-response bias, program administrators should keep track of which potential respondents have returned surveys and which have not, and send notes of reminder or even a second (and possibly third) wave of surveys. Monitoring respondents and non-respondents will also allow for a determination of response rates.

To avoid respondent fatigue and increase response rates, questionnaires should be constructed so that respondents can complete them in a reasonable period of time (e.g., 10-20 minutes). The use of secure and confidential interactive computer surveys should also be considered.

Guideline 6-7. Questionnaire content and wording should be structured with the relevant respondent group, and the nature and extent of that group's interaction with judges, in mind. In most instances, it will be necessary to use a different performance questionnaire for each respondent group.

Guideline 6-8. Judges should be evaluated periodically. The frequency of judicial evaluations should be related to such factors as the length of time the judge has served on the bench and when the judge will be considered for reappointment, retention, or reelection.

Commentary

In order to meet the goal of improving judicial performance, the results of judicial performance evaluation must be shared with the judge on a regular and periodic basis. The frequency of evaluations of similarly situated judges should be consistent.

New judges should have the benefit of receiving performance evaluation results early in their judicial career. At the same time, a new judge may require some time to become accustomed to serving on the bench, and a premature evaluation may be unfair and unproductive. To strike a balance between these two concerns, it is recommended that evaluations occur two years after judicial service commences.

More experienced judges may be evaluated with less frequency. The primary purpose of evaluating such judges is to help them assess whether they have improved their performance since the last evaluation. While it would certainly be appropriate to evaluate such judges with greater frequency, four years is the maximum time that should transpire between evaluations.

When performance evaluations are used to enable those responsible for continuing judges in office to make informed decisions, such evaluations should, at a minimum, coincide with the times at which such decisions are to be made.

SELECTED REFERENCES ON JUDICIAL EVALUATION

Judicial Performance Evaluation in the States

American Bar Association National Conference of State Trial Judges. 1996. *Judicial Performance Evaluation Handbook*. Chicago: American Bar Association.

Greenstein, Marla N., Dan Hall, and Jane Howell. 2002. "Improving the Judiciary through Performance Evaluations." In Gordon M. Griller, E. Keith Stott, and John Fallahay, eds., *The Improvement of the Administration of Justice*. Chicago: Lawyers Conference, Judicial Division, American Bar Association.

Kearney, Richard C. 1999. "Judicial Performance Evaluation in the States." *Public Administration Quarterly* 22:468.

Keilitz, Susan, and Judith White McBride. 1992. "Judicial Performance Evaluation Comes of Age." *State Court Journal* 16:4.

Judicial Retention Evaluation Programs

Andersen, Seth S. 2001. "Judicial Retention Evaluation Programs." *Loyola of Los Angeles Law Review* 34:1375.

Esterling, Kevin M. 1999. "Judicial Accountability the Right Way." *Judicature* 82:206.

Esterling, Kevin M., and Kathleen M. Sampson. 1998. *Judicial Retention Evaluation Programs in Four States: A Report with Recommendations*. Chicago: American Judicature Society.

White, Penny J. 2002. "Judging Judges: Securing Judicial Independence by Use of Judicial Performance Evaluations." *Fordham Urban Law Journal* 29:1053.

Methodology

Bernick, E. Lee, and David J. Pratto. 1995. "A Behavior-Based Evaluation Instrument for Judges." *Justice System Journal* 18:173.

Brody, David C. 2000. "Judicial Performance Evaluations by State Governments: Informing the Public While Avoiding the Pitfalls." *Justice System Journal* 21:333.

Brody, David C. 2004. "The Relationship Between Judicial Performance Evaluations and Judicial Elections." *Judicature* 87:168.

Durham, Christine M. 2000. "Gender and Professional Identity: Unexplored Issues in Judicial Performance Evaluation." *The Judges' Journal* 39:11.

Farthing-Capowich, Daina. 1985. "Designing Programs to Evaluate Judicial Performance: Participation Is Key to Success." *State Court Journal* 9:22.

Farthing-Capowich, Daina, and Judith White-McBride. 1987. "Obtaining Reliable Information: A Guide to Questionnaire Development for Judicial Performance Evaluation Programs." *State Court Journal* 11:5.

Sterling, Joyce S. 1993. "The Impact of Gender Bias on Judging: Survey of Attitudes Toward Women Judges." *Colorado Lawyer* 22:257.