

# 9

## Judicial Performance Evaluation

Malia Reddick and Hon. Rebecca Love Kourlis

While the system enshrined in the U.S. Constitution for selecting and retaining federal judges preserves their independence from majority will and other political pressures, most processes for selecting state court judges build in a mechanism for holding judges accountable for their performance on the bench, whether through reappointment, retention election, or reelection. In the mid-1970s, states began to undertake the development of judicial performance evaluation programs that would ensure this judicial accountability is meaningful—that it is driven by impartial and reliable information from a wide range of sources and is based on the process rather than the outcomes of judging.

Judicial performance evaluation (JPE) offers a number of benefits to state judiciaries and the public. First and foremost, JPE programs provide a valuable source of information to voters and other who reselect

---

Malia Reddick is manager of the Quality Judges Initiative at IAALS—the Institute for the Advancement of the American Legal System at the University of Denver. She is a nationally recognized expert in the areas of judicial selection and judicial performance evaluation. She earned her PhD in political science from Michigan State University.

Rebecca Love Kourlis is founder and executive director of the Institute for the Advancement of the American Legal System (IAALS) at the University of Denver. She is a former trial court judge and served for 12 years as a Colorado Supreme Court justice. She received her JD from Stanford Law School.

judges about the on-the-job performance of jurists who are seeking additional terms. In many states where judges face retention elections, JPE programs may provide the only substantive, objective source of information about judges on the ballot.

JPE serves a broader purpose as well in the reselection context, in that it educates the public about the specific qualities and behaviors that make a good judge. With ever-increasing frequency and fervency, courts and judges are assessed according to the political popularity of their decisions. When judges make sound legal rulings with which critics disagree ideologically, they are often labeled as “judicial activists” who “legislate from the bench.” They are portrayed as policy makers rather than neutral arbiters of the law. JPE programs can counteract these messages by educating citizens about the essential qualities of state court judges, including fairness, legal ability, strong communication skills, and professionalism. In this regard, JPE helps to preserve judicial independence.

Of equal value is the fact that JPE provides judges with the necessary tools to improve their own performance. The feedback garnered through JPE enables judges to identify their strengths and weaknesses and can motivate them to capitalize on the strengths and mitigate the weaknesses. In addition, chief judges may use evaluation results to foster professional development among the judges they supervise, and the judicial branch may use JPE findings to develop or enhance continuing education programs for judges.

In addition to these primary purposes that JPE serves, such programs have the additional benefit of enhancing public trust and confidence in the judiciary by demonstrating that individual judges and the judiciary as a whole are accountable for their performance.

Many states have “official” programs for evaluating judicial performance that are authorized by constitution, statute, or court rule, and designed and implemented by state entities. “Unofficial” JPE programs are also in place around the country, sponsored by bar associations, civic organizations, the media, or special interest groups.

Official JPE programs are most commonly found in states where judges are selected through a commission-based gubernatorial appointment process, commonly known as merit selection or the Missouri Plan. JPE is also one of four components of the O’Connor Judicial Selection Plan, endorsed by former U.S. Supreme Court Justice Sandra Day O’Connor to ensure impartiality, accountability, and transparency in choosing judges. Six states—Alaska, Arizona, Colorado, Missouri, New Mexico, and Utah—use such a process to select at least some judges. In

these states, JPE provides broad-based, apolitical information to voters about judges standing for retention.

State-run programs for evaluating judges' performance are not confined to states with judicial retention elections. JPE serves a similar purpose in states like Connecticut and Virginia, in which the legislature plays a role in reappointing judges. Trial judge performance evaluations factor into legislators' decisions to reappoint those judges to another term. In the District of Columbia and Hawaii, the judicial nominating commissions that screen judicial applicants and recommend the best qualified for appointment by the governor also decide whether to reappoint judges, with the benefit of JPE reports. And in those states where judges have life tenure—Massachusetts, New Hampshire, and Rhode Island—official JPE programs are used to encourage and inform judicial self-improvement.

These official programs differ in their details—the specific criteria on which judges' performance is assessed, the sources of the feedback that is sought, and the extent of information that is made public—but they share a common focus on the process, rather than the outcomes, of judging.

Unofficial JPE programs tend to emphasize the voter information objective of JPE. State bar associations in a number of states have programs where attorneys evaluate or rate judges standing for retention or reelection; many county bar associations conduct the same type of program for local trial judges. Some bar associations offer preference polls, where attorneys simply indicate whether they would retain the judge in office, while others ask attorneys to gauge the extent to which judges are qualified to be retained. Some bar evaluation processes go further by asking attorneys to rate judges on several performance-based criteria.

Civic organizations in some states may offer JPEs as a tool for voters. For example, in 2014, a group of attorneys, legal academics, court professionals, and citizens in Kansas known as the Judicial Performance Evaluation Committee formed a coalition to evaluate appellate justices standing for retention. The process was modeled after the state's official JPE program that was defunded by the legislature in 2012.

Media evaluations of judges were more common in the past but have dwindled out in recent years. One such program remains active, however. Since 1992, the *Las Vegas Review-Journal* has offered "Judging the Judges," a biannual evaluation of supreme court justices and trial court judges in which Clark County attorneys rate judges' adequacy on a range of qualities and indicate whether they recommend each judge be retained in office.

Unofficial JPE programs offered by special interest groups typically present a one-sided and incomplete perspective of judges' work, basing the performance assessment solely on how judges have decided cases involving hot-button issues like same-sex marriage, abortion rights, tort reform, capital punishment, or taxation. Such programs are often conducted in conjunction with appellate judges standing for retention or reelection, and may be created in the first instance for this purpose.

## ORIGINS OF JPE

Straw polls conducted by state and local bar associations represent the earliest efforts to gauge court users' opinions of judges' performance. These polls indicated lawyers' preferences as to whether judges should be retained in office, whether by retention election, reelection, or reappointment. Over time, straw polls evolved into attorney evaluations of judges' on-the-job performance according to set criteria.

Members of the judiciary, the legal profession, and the public voiced concerns about these attorney evaluations. First and foremost, the primary purpose of the evaluations was to inform the reselection decision, and little attention was paid to developing surveys whose results would facilitate self-improvement. In addition, to provide useful information to voters, the surveys needed to move beyond attorneys as the sole respondents and include other groups who had contact with judges in the courtroom; and, even among attorneys, response rates were often low. Finally, to the extent bar evaluations were a useful source of information to those responsible for reselecting judges, they were not widely publicized outside the legal community.

In the late 1970s, a handful of entities began to experiment with more broadly based evaluation processes that would also provide the feedback necessary to improve the quality of judges' work on the bench. The state of Alaska led the way in 1974, with the Alaska Judicial Council conducting attorney surveys and soliciting public comment on judges standing for retention.<sup>1</sup> Alaska established an official JPE program in 1975. It was a serious first-time effort, with the council recommending against one judge's retention.

At the national level, the topic of a more comprehensive JPE program was first raised at an American Bar Association (ABA) meeting in 1978, and

1. E-mail from Susanne DiPietro, Exec. Dir., Alaska Judicial Council, to author (Nov. 5, 2014, 21:56 MST) (on file with author).

the ABA formed the Evaluation of Judicial Performance Committee in 1979. This committee published *Black Letter Guidelines for the Evaluation of Judicial Performance* in 1985. Also in 1979, the National Center for State Courts (NCSC) drafted an internal concept paper calling for further exploration of this innovation. That same year, the New Jersey Supreme Court Committee on Judicial Performance and Evaluation recommended the creation of an ongoing JPE program. A pilot program was conducted in New Jersey in 1983, and formal rules for an official process were adopted in 1988.

JPE by state judicial branch entities took off in the 1980s, with six states—Connecticut, Utah, New Hampshire, Illinois, Colorado, and New Jersey—establishing programs to better equip voters in judicial retention elections, promote the improvement of judges' performance, or both. The trend continued in the 1990s, with Arizona, Hawaii, Rhode Island, New Mexico, Florida, and Idaho adopting JPE. Massachusetts was close on these states' heels, enacting its program in 2001. As of 2015, 17 states and the District of Columbia offer official programs for evaluating judicial performance.<sup>2</sup>

## KEY COMPONENTS OF JPE PROGRAMS

The 18 official JPE programs around the country, as well as the state bar evaluation processes in place in at least 15 states, vary fairly extensively on the particulars. However, we can identify some key components among both state- and bar-sponsored programs, and we discuss these below.

### Evaluation Commission

In most states with an official JPE program, and in all states where the JPE program provides voter information for retention elections, the program is overseen and administered by a commission created for this purpose.<sup>3</sup> These commissions are typically composed of attorneys, members of the public, and often, sitting or retired judges. Commission members may be appointed by a variety of government officials, including the governor, legislative leaders, the supreme court, and the state bar association's board of governors. Members serve staggered terms of fixed length.

---

2. JPE programs in the District of Columbia and Vermont are not standalone programs but are incorporated in the process for reselecting judges. In 2008, the Missouri Supreme Court adopted a rule expanding the Missouri Bar retention evaluation poll that had been conducted since 1948.

3. Alaska is an exception. A constitutionally authorized body known as the Alaska Judicial Council administers the JPE program in addition to other responsibilities.

## Evaluation Criteria

The crux of any JPE program is the criteria on which judges are assessed—the principles and standards to which they are held. One contribution of the original bar polls was to identify commonly accepted criteria on which judges' performance should be evaluated, at least from the attorney perspective. Today, virtually all performance evaluation programs, whether state- or bar-sponsored, employ criteria that mirror those recommended in the 2005 American Bar Association Guidelines for the Evaluation of Judicial Performance: legal ability, integrity and impartiality, communications skills, professionalism and temperament, and administrative capacity. Some JPE programs also assess diligence and service to the legal profession and the public.

## Evaluative Tools

Whether conducted as part of a state-run program or by a state bar association, attorney surveys are the primary tool used to evaluate judicial performance. Most official JPE programs only send surveys to those attorneys who have appeared before a judge during the period of evaluation. In bar programs, on the other hand, it is common for all licensed attorneys in the state to receive the survey, with an admonishment to only complete the survey if they have had recent professional contact with the judge. One advantage of surveying electronically is that attorneys may be barred from completing the survey if their answers to threshold questions indicate a lack of such contact.

In the interest of a broad-based performance evaluation, official JPE programs survey other groups that interact professionally with judges as well, including court staff, other judges in the state, and jurors. Some programs also survey law professors, litigants, witnesses, and state employees who regularly appear in court, such as peace and probation officers and social workers.

With all respondents, those who administer the survey follow strict procedures to ensure anonymity. Where respondents provide narrative comments to be shared with evaluated judges, JPE commission staff members take special care to conceal the identity of the author.

The courtroom observation programs used in some states, particularly for trial judges, provide an opportunity for "real time" assessments of on-the-bench performance. Commission members themselves or citizen

volunteers conduct these observations.<sup>4</sup> Nearly all of the official JPE programs for retention purposes also invite broader public input, whether through public hearings or the submission of written comments.

JPE programs may also examine the outputs, whether direct or indirect, of the judge's work. For appellate judges, the primary work product is the written opinion, so a process for assessing the quality and clarity of these opinions is an essential component of a performance review program. Most official appellate JPE programs incorporate written opinion review, though there is wide variation in how these processes are structured and carried out.<sup>5</sup> Court records may also be examined as objective measures of both trial and appellate judges' performance. Examples of such records that official JPE programs take into account include case management data, reversal rates on appeal, recusal and peremptory challenge records, and compliance with disciplinary, education, and case-under-advisement time standards.

Many programs also invite judges to assess their own performance on the bench, via a self-evaluation survey, so that they may see how their own views line up with those of court users and colleagues. Several official programs require an in-person interview with each judge, to discuss evaluation results, potential concerns, and plans for addressing any concerns; the judge's self-evaluation survey may be one focus of the interview.

## Public Narrative and Retention Recommendation

When one of the purposes of a JPE program is to inform voters, the JPE commission prepares a narrative summary for each judge of the results of the surveys and other evaluation tools. Narratives typically also include a photo and bio of the judge and a retention recommendation. JPE commission members decide, usually by vote, what recommendation to make to voters as to each judge's retention. This may take the form of a "Retain/Do Not Retain" determination or an indication that a

4. See Nicholas H. Woolf & Jennifer MJ Yim, *The Courtroom-Observation Program of the Utah Judicial Performance Evaluation Commission*, 47 CT. REV. 84 (2011) (describing an innovative courtroom observation program).

5. IAALS offers guidelines and tools for an opinion review process. Inst. for the Advancement of the Am. Legal Sys., *Recommended Tools for Evaluating Appellate Judges* (2013), <http://iaals.du.edu/initiatives/quality-judges-initiative/recommended-models/recommended-tools-for-evaluating-appellate-judges>.

judge “Meets/Does Not Meet Judicial Performance Standards.” In some instances, a commission may decide that it does not have enough information to make a recommendation. In bar-sponsored programs, the bar association may publicize the percentage of attorneys who support each judge’s retention, or whether attorneys recommend a judge for retention or find a judge qualified for reelection.

Where a JPE commission makes a recommendation as to whether judges should be retained in office, judges are given an opportunity to meet with the commission and respond to the narrative summary and retention recommendation. They may also be permitted to submit a response in writing. The commission may make changes to the narrative and/or the recommendation at its discretion.

### **Self-Improvement Mechanism**

State-sponsored JPE programs share a common objective—to provide judges with the feedback necessary to facilitate self-improvement. Many of these programs explicitly build in a mechanism to accomplish this, often in cooperation with the state’s judicial education entity. The process may range from developing individualized improvement plans for judges to designing continuing education programs to address common issues.<sup>6</sup>

As one of the primary purposes of many JPE programs is to provide performance assessments to those who reselect judges, evaluations are typically conducted at the end of a judge’s current term of office. Most official JPE programs also conduct an interim evaluation at least once during each judge’s term to facilitate self-improvement.

## **JPE AROUND THE COUNTRY**

To provide a sense of how these components come together in functioning JPE programs, we offer several examples of official and state bar association programs in place around the country that are used to promote the various purposes of JPE, and we describe these programs in some detail below. Note that while we identify a primary purpose for

---

6. Arizona’s Judicial Performance Review program uses a conference team approach for judicial self-improvement that is unique to that state. Judicial Performance Review, JPR Conference Teams, <http://www.azcourts.gov/jpr/JPRConferenceTeams.aspx> (last visited Jan. 12, 2015).



which each program was established—informing voters and others who select or reselect judges, facilitating judicial self-improvement, and fostering public confidence in the judiciary—these programs serve additional purposes as well, whether intentionally or unintentionally.

## Official Programs

### *Informing retention election voters in Colorado*

Colorado judges are chosen through a commission-based gubernatorial appointment process, with judges standing in yes/no retention elections for subsequent terms. Term lengths range from four to ten years, depending on the level of court. In 1988, following two election cycles with successful pilot JPE programs administered by the Colorado Judicial Institute, the state legislature established a JPE program.<sup>7</sup> The program's purpose is twofold: "to provide persons voting on the retention of justices and judges with fair, responsible, and constructive information about judicial performance and to provide justices and judges with useful information concerning their own performances."<sup>8</sup> To achieve these purposes, judges are evaluated midterm and prior to standing for retention.

The state commission on judicial performance administers evaluations of appellate judges and oversees 22 local district commissions that evaluate trial judges in each judicial district and county. Each commission has ten members, comprising six nonattorneys and four attorneys. The governor and the chief justice each appoint one attorney member and two nonattorney members, and the speaker of the house and the president of the senate each appoint one attorney member and one nonattorney member. Commission members serve staggered, four-year terms.

All judges are evaluated on six criteria, including integrity, legal knowledge, communication skills, judicial temperament, administrative performance, and service to the legal profession and the public. For appellate judges, surveys are sent to attorneys, court staff, other appellate judges, and district judges. For trial judges, surveys are sent to attorneys, jurors, witnesses, litigants, court staff, court interpreters, law enforcement personnel, crime victims, social service caseworkers, and appellate

7. The Colorado Judicial Institute is a nonpartisan, nonprofit organization of lawyers and citizens dedicated to excellence in Colorado's judicial system.

8. COLO. REV. STAT. § 13-5.5-101(1) (2014).

judges. However, surveys are only one component of the evaluation process. Commissions also take into account judges' self-evaluations, courtroom observation by commission members, case management data, interviews with judges, and review of judges' written decisions or opinions.

For each judge standing for retention, the appropriate state or district commission compiles a narrative that includes a brief biography, a summary of the survey results (with a link to full survey results), and an overview of the judge's performance with respect to the other bases for evaluation. The commission also makes a recommendation as to whether the judge should be retained. These narratives are included in a voter guide mailed to all households (the Colorado Blue Book) and posted on the website of the office of judicial performance evaluation (OJPE). Though conducted primarily for the purpose of judicial self-improvement, survey results for midterm evaluations are also posted on the OJPE's website.

### ***Reappointing judges in Connecticut***

Connecticut's JPE program for superior court judges was first authorized by directive of the chief justice in 1984. It was designed as a self-improvement tool for judges to enhance their performance on the bench. Over the years, the program evolved and is now used not only by the judges for their own improvement, but as a tool for the judicial selection commission and state legislature in the judicial reappointment/elevation process. Superior court judges are chosen through a commission-based gubernatorial appointment process with legislative confirmation.<sup>9</sup> Judges serve eight-year terms and may be reappointed. To inform the reappointment decision, JPE results are provided to the state judicial selection commission, which evaluates judges seeking reappointment to the same court and sends names of those recommended for reappointment to the governor, and to the general assembly's joint standing committee on judiciary, which decides whether to confirm judges nominated by the governor.

The JPE program also fosters judicial self-improvement. Composite survey results are shared with individual judges and with the chief court administrator who reviews results with individual judges.

---

9. According to Article 5 and Amendment Article XX of the Connecticut Constitution, judges "shall, upon nomination by the governor, be appointed by the general assembly." CONN. CONST. art. V, § 2, *amended by* CONN. CONST. amend. XX.

The preparation of composite survey results and the review of results with individual judges occurs after a minimum of 25 completed attorney surveys have been returned to the program administrator and a minimum of two years have elapsed since the preparation and review of the previous composite survey results.

The JPE program in Connecticut is entirely survey-based. Attorneys are asked to rate superior court judges as excellent, good, fair, or poor through a series of questions in the categories of comportment, legal ability, and management skills. Jurors are given an opportunity to rate judges on issues of equality and fairness through a series of “attitude toward” questions and on other topics including dignity of proceedings, attentiveness, patience, courtesy, explanation of proceedings, efficiency, and clarity of charge.

Survey forms are distributed to attorneys and jurors on an ongoing basis. They are distributed to attorneys who appeared before judges in full trials, partial trials in which the case settled after evidence was presented, mistrials, and hearings lasting one hour or longer. Surveys are also distributed to jurors who entered the deliberation stage of a trial or who participated in at least two days of a trial.

In 2012, court administrators launched an electronic survey for judges assigned to high-volume courts, which calls for semiannual surveying of eligible attorneys who have appeared before judges in certain types of criminal matters.

### ***Improving individual judicial performance in Illinois***

The Illinois Supreme Court has offered a voluntary JPE program for circuit and associate judges since 1988, but in 2011, a new supreme court rule made the program mandatory. According to Rule 58, the program is designed “for the purpose of achieving excellence in the performance of individual judges and the improvement of the judiciary as a whole.” The supreme court’s JPE committee, a 21-member body composed of judges, attorneys, and academics, monitors the evaluation program.

Judges are randomly selected for evaluation each year and are asked to identify attorneys and court personnel to be surveyed. Evaluation criteria include legal reasoning and ability, impartiality, professionalism, communication skills, and management skills. Evaluation results are strictly confidential and shared only with the evaluated judge and a facilitator, who is a judge or retired judge trained in the performance

improvement process. Following an initial evaluation, subsequent evaluations are voluntary on the part of the judge.

### ***Promoting public confidence in the courts in New Hampshire***

The New Hampshire Judicial Branch has evaluated its judges since 1987. Starting in 2001, trial judges, marital masters, and hearing officers have been evaluated every three years. Supreme court evaluations began in 2008.

For trial judges, surveys are made available to attorneys, litigants, court staff, state and local agencies, and jurors, who rate judges on a five-point Likert scale. The questions focus on such criteria as degree of preparedness, attentiveness, objectivity, legal knowledge, performance (including the ability to identify and analyze issues, and application of the law), judicial management skills (including making the juror's role clear), and the thoroughness and timeliness of decisions.

For the supreme court, surveys are sent every three years to attorneys, litigants, law professors, judges, and marital masters, who evaluate them in a variety of performance areas, based on a five-point Likert scale.

The performance evaluation process is primarily survey-based, but additional evaluation tools are utilized. Each year, the administrative office of the courts analyzes the supreme court's processing of cases to determine whether objective time standards were met. The justices themselves meet to assess each other's performance in this regard. Justices and judges also complete self-evaluation questionnaires.

The JPE committee compiles an annual report on the JPE program that includes summary JPE results for each level of court. The report is available to the public on the judicial branch website and provided to the governor and legislative leadership as required by state law.

## **State Bar Association Programs**

### ***Informing Retention Election Voters in Iowa***

The Iowa State Bar Association (ISBA) has conducted performance reviews of judges since the state adopted its commission-based gubernatorial appointment process in 1962. Judges serve terms of six to eight years, depending on the level of court, and then stand in a yes/no retention election. The ISBA offers attorney ratings of judges as a source of voter information.

Attorneys who have appeared before judges standing for retention are asked to rate each judge in several areas related to their professional competence, with ratings on a 1–5 scale ranging from “very poor” to “excellent.” Attorneys also rate each judge in areas related to their demeanor, with ratings on a 1–5 scale ranging from “strongly disagree” to “strongly agree.” Evaluation results are posted on the ISBA website.

### ***Informing Voters in Contestable Elections in North Carolina***

North Carolina chooses its judges in nonpartisan elections. While state and local bar associations around the country offer ratings of incumbent justices and judges seeking reelection, the North Carolina Bar Association (NCBA) is one of only a few state bar associations that evaluates both sitting judges and their attorney challengers.<sup>10</sup> Following a series of pilot programs conducted from 2007 to 2010, the NCBA launched a full JPE program in 2011 to provide voters with information about trial judge candidates in the 2012 election cycle.

Survey-based performance ratings for all judicial candidates are made available to the public and the media for the primary and general elections. For incumbent superior and district court judges, attorneys who have had sufficient professional contact with trial judges are asked to rate them on criteria that include integrity and impartiality, legal ability, professionalism, communication, administrative skills, and overall performance. Attorneys who have had sufficient professional contact with nonjudge candidates are asked to rate them on the same six criteria. Ratings are on a five-point scale ranging from “poor” to “excellent.”

Surveys for sitting judges are conducted in October of odd-numbered years, and surveys for nonincumbent challengers are conducted in March of each election year. Survey results for sitting judges are made available in January of each election year, well before the candidate-filing period closes, and results of the attorney challenger surveys are released in April, before the judicial primary election takes place in May.

In 2013, the NCBA offered a voluntary, confidential evaluation program for new judges, the rationale for which is to allow judges to go through an initial assessment process prior to the pre-election survey. The process begins with the chief judge in each district identifying 10–15 attorneys who have had professional contact with each new

---

10. The NCBA is a voluntary state bar association.

judge. Then, retired judges and attorneys who have volunteered to assist with the evaluation program interview these attorneys regarding each judge's performance and prepare a confidential written report for the judge.

### ***Reappointing judges in Maine***

Since 1997, the Maine State Bar Association (MSBA) has evaluated the performance of state trial judges, for the purpose of making a recommendation as to whether judges should be reappointed. Maine judges are nominated by the governor and confirmed by the senate for seven-year terms and may be renominated and reconfirmed.

Trial judges are evaluated as late as possible in the sixth year of each term. Evaluations are survey-based, with all Maine attorneys invited to rate judges on qualities related to legal ability, integrity, impartiality, judicial temperament, diligence, and special skills. The MSBA's judicial evaluation committee prepares a report on each judge for the MSBA's board of governors that includes a concise summary of survey results, a recommendation as to whether each judge is qualified for renomination, and the reasons for such a recommendation. The board of governors reviews the report to determine its own recommendation as to renomination, which is then shared with the judge, the chief justice, and the governor.

The MSBA also conducts confidential evaluations during the second year of a judge's initial term to allow the judge to assess his or her performance. All evaluation results are also made available to each judge's chief judge and to the chief justice of the Maine Supreme Judicial Court.

## **IMPACT OF JPE**

The impact of JPE, as embodied in the programs described above, must be measured according to the extent to which it accomplishes its chief intended purposes—providing objective and meaningful information to those who make reselection decisions and giving judges the necessary feedback to improve their own performance. Published in 1998, the American Judicature Society (AJS) conducted the first multistate, comprehensive study of the effectiveness of JPE.

In the mid-1990s, AJS undertook an examination of judicial retention evaluation programs in the four states in which they were in

place—Alaska, Arizona, Colorado, and Utah.<sup>11</sup> One component of the study explored the extent to which JPE reports for judges on the ballot influenced voter decisions, and researchers approached this question in two ways. In exit surveys of voters who had received evaluation information, at least three-fifths of respondents in each state reported that the information influenced their retention votes.<sup>12</sup> Approximately two-thirds of those familiar with JPE reports in each state said the reports made them more likely to vote in judicial elections.<sup>13</sup> Researchers also examined whether voters found judges' evaluation results persuasive, based on an analysis of election data. Using two decades of election results from Alaska, they found a statistical correlation between a judge's numerical rating and his or her affirmative vote percentage in the retention election—the higher a judge's score, the higher his or her share of "yes" votes.

The AJS study surveyed judges themselves in these four states, as well, regarding their views on the value of JPE. A supermajority of judges in each state, ranging from 74 percent in Utah to 85 percent in Arizona, believed that their JPE reports provided useful feedback on their performance. In addition, three-quarters of the judges in Alaska, Arizona, and Colorado reported that the evaluation process held them appropriately accountable for their performance, and 59 percent of Utah judges said the same.

Most subsequent analyses of the effectiveness of JPE have utilized the AJS study as a template. We summarize some of these state-specific studies here.

## JPE as Voter Information

A follow-up study of retention ratings and votes in Alaska, based on data from the 2000 to 2006 elections, confirmed that higher ratings were associated with a higher percentage of "yes" votes.<sup>14</sup> A review of Arizona's 2012 retention election data also showed a correlation between

11. KEVIN M. ESTERLING & KATHLEEN M. SAMPSON, AM. JUDICATURE SOC'Y, JUDICIAL RETENTION EVALUATION PROGRAMS IN FOUR STATES: A REPORT WITH RECOMMENDATIONS (1998).

12. The exit surveys were conducted in each state's largest locality.

13. AJS researchers were careful to note that these figures may overstate the impact of JPE on voters, as respondents to the exit survey were a nonrandom group selected because of their knowledge of the retention evaluation reports.

14. David C. Brody, *The Use of Judicial Performance Evaluation to Enhance Judicial Accountability, Judicial Independence, and Public Trust*, 86 U. DENV. L. REV. 1 (2008).

JPE results and retention votes. As the number of commissioners who rated a judge negatively increased, so did the percentage of “no” retention votes for the judge.<sup>15</sup>

While some evidence suggests that voters take judicial evaluation results into account, a negative retention recommendation does not necessarily lead to a judge’s ouster. Only a small percentage of judges receive a “do not retain” recommendation. For instance, since 1974, the Alaska Judicial Council has recommended that only 12 judges not be retained. This fact may arguably be attributed to the success of commission-based appointment systems in choosing high-quality judges, as they are designed to do. But even those judges who do receive poor evaluations are not always voted out of office. For example, in the 2014 election cycle, official JPE commissions in seven states recommended against the retention of eight judges, and voters retained all but three of these judges.

It is rare, in any event, for judges not to be retained. A study of judicial retention elections in ten states from 1964 to 2006 revealed that only 56 of 6,306 judges lost their bids for retention, with ten of the nonretained judges serving in Illinois where the retention threshold is 60 percent.<sup>16</sup>

However, though retention outcomes do not always correlate with JPE ratings or recommendations, this is not the full measure of the efficacy of JPE programs in maintaining a high-quality judiciary. JPE commissions share evaluation results with sitting judges well in advance of the filing deadline for retention elections. Many judges who receive unfavorable evaluations and/or negative recommendations choose not to stand for election, whether to avoid professional embarrassment or defeat at the polls. Thus JPE is effective—albeit often indirectly—as an accountability mechanism.

Some scholars have approached the larger question of whether voters are aware of judicial performance evaluations in an indirect manner, by exploring the impact of JPE on voter roll-off in judicial elections. “Roll-off” refers to voters who participate in top-of-the-ballot races but do not cast votes in lower-ballot contests like judicial elections. The working hypothesis is that voters already at the polls are more likely to vote in judicial races when they have information about the candidates, but supporting evidence for this hypothesis is anecdotal at best. For example,

15. Rebecca White Berch & Erin Norris Bass, *Judicial Performance Review in Arizona: A Critical Assessment*, 56 ARIZ. L. REV. 353 (2014).

16. Larry Aspin, *Judicial Retention Election Trends 1964–2006*, 90 JUDICATURE 208 (2007).



one scholar examined voter roll-off in judicial elections in Arizona, Colorado, and New Mexico after JPE was adopted.<sup>17</sup> Results were inconclusive, with roll-off in these races declining in Colorado and New Mexico but increasing in Arizona. More comprehensive research on what voters know in judicial retention elections—and how they know it—is essential to helping JPE reach one of its most crucial target audiences.

### JPE as a Self-Improvement Tool

The most common approach taken to gauging whether JPE facilitates judicial self-improvement is asking judges themselves. The Institute for the Advancement of the American Legal System (IAALS) did just that in 2008, conducting a survey of Colorado judges regarding their perceptions of the state's performance evaluation process.<sup>18</sup> More than 85 percent of trial judges, and half of appellate judges, said that JPE had been beneficial to their professional development. More than 70 percent of judges who had been evaluated at least once reported that the process provided them with feedback that enabled them to improve their job performance.

When asked about the effect of the JPE process on their judicial independence, the most common response from both appellate and trial judges was that the process had no effect, while approximately one-fourth of judges said that JPE *increases* their judicial independence. In addition, although judges were not surveyed regarding the impact of JPE on their accountability, several judges mentioned in their written comments that JPE made them more accountable. According to one judge, it helps to stave off “black robe disease”; another described JPE as “a good hedge against judicial arrogance.”

To provide a foundation for discussions at a 2011 national conference on evaluating appellate judges, IAALS conducted another judicial survey, this time of appellate judges in nine states who were subject to JPE. Regarding the impact of JPE on their professional development, 53 percent of respondent judges found it “somewhat beneficial” and 10 percent viewed it as “significantly beneficial.”

17. See Brody, *supra* note 14.

18. INST. FOR THE ADVANCEMENT OF THE AM. LEGAL SYS., THE BENCH SPEAKS ON JUDICIAL PERFORMANCE EVALUATION: A SURVEY OF COLORADO JUDGES (2008), [http://iaals.du.edu/images/wygwam/documents/publications/Bench\\_Speaks\\_On\\_JPE2008.pdf](http://iaals.du.edu/images/wygwam/documents/publications/Bench_Speaks_On_JPE2008.pdf).

## AREAS OF CONCERN

While there is some empirical evidence that JPE programs are accomplishing their stated objectives, there are also some potential concerns that, if not addressed and resolved, may impair the full achievement of these programs' common goals. We discuss two of these below—low survey response rates and the potential for implicit biases to affect JPE.

### Low Response Rates

One of the early criticisms of bar association polls—low response rates—is an issue that has plagued official JPE programs as well. Researchers have raised concerns about unrepresentative data in general, whether as a result of coverage error, where a portion of the desired sample is inaccessible; sampling error, where only a portion of the desired sample is surveyed; and/or nonresponse error, where members of the desired sample opt not to respond to the survey.<sup>19</sup>

Official JPE programs go to great lengths, as well great expense, to minimize these data problems to the extent possible, while still providing essential, honest assessments to voters of judges' on-the-job performance and giving judges practical feedback about their strengths and weaknesses on the bench. Program administrators recognize as well the significant professional implications that evaluation results can have for judges, and they take their responsibility in this regard very seriously.

Most JPE programs consult with independent entities such as research and polling firms or university research institutes to handle the collecting and processing of survey responses from the various respondent groups. Consider, for example, the surveying process in two official JPE states that use the same outside contractor.

The contractor receives case data from a range of primary court sources with the names of all attorneys who had likely been in each judge's courtroom. Attorneys who appeared before judges up for evaluation in a given year receive an initial letter inviting them to complete the survey online and providing a link to the survey with a unique password. Reminder e-mails are sent for two consecutive weeks, and if attorney respondents still do not complete the survey, they receive a phone call with a final invitation to complete the survey by phone or online.

<sup>19</sup> See, e.g., Jennifer K. Elele, David B. Rottman & Brian L. Cutler, *Judicial Performance Evaluation: Steps to Improve Survey Process and Measurement*, 96 JUDICATURE 65 (2012).

Nonattorneys with courtroom experience with the evaluated judge (e.g., court employees, jurors, probation officers, witnesses) are identified using the same primary court sources. (The only exception is criminal defendants and civil litigants, of whom a random sample is taken.) Potential nonattorney respondents with known e-mail addresses receive e-mails and follow-up e-mails inviting them to complete the survey online. Other nonattorney respondents receive two waves of letters with postage-paid return envelopes.

Even in this thorough and meticulous process, data problems may occur. A judge's survey response rates may be low, his or her evaluation scores may be skewed by extreme responses at either end of the spectrum, or survey respondents may be unrepresentative of those with whom the judge has interacted. Fortunately, there are some fallbacks in place to guard against these data problems having negative ramifications for judges.

It is becoming increasingly common for JPE surveys to include questions about respondent demographics, including the nature and extent of the respondent's interaction with the judge, the nature and extent of the attorney respondent's legal experience, and the subject matter of the case(s) in which the respondent was involved, if any, and its outcome. These statistics can be viewed alongside the judge's survey results, and both the judge and the commission can assess the representativeness of the survey respondents. Together they can make a decision as to whether it is fair to include the results in the judge's evaluation. As well, in official JPE programs there are other evaluative tools in addition to the survey results upon which the commission may rely for a complete picture of the judge's performance, including courtroom observation, case management data, written decisions or opinions, personal interviews, and public input.

In short, JPE in most states is more than simply a "data in, data out" process. An important component of the JPE commission's role is to monitor the entire evaluation effort and exercise discretion and judgment as necessary to ensure the fairness and accuracy of the information that is made public.

### **Potential for Implicit Biases to Affect JPE**

A more recent concern that has arisen in the context of JPE is the potential for implicit biases based on gender, race, or ethnicity to affect evaluations of judges. More specifically, the fear is that survey respondents

may systematically score women or minority judges lower than white male judges in JPE surveys because of unconscious expectations of how individuals should behave based on their gender or race. One of the negative ramifications of these biases is that women and minority judges are placed at a disadvantage in the eyes of those responsible for retaining or reappointing them.

On their face, a few empirical studies seem to validate this concern. An examination of the process used by state and city bar associations in Missouri from 1998 to 2006 to assess judicial performance and inform voters revealed that male and Caucasian judges were recommended for retention at a significantly higher rate than were women judges and judges of color.<sup>20</sup> Of all demographic groups, women judges of color were least likely to be recommended for retention. The author of the study discussed the possibility that these disparities were due to implicit biases on the part of survey respondents, but also recognized that they could be a product of flaws in the evaluation process, real differences in performance, or a combination of these factors.

An analysis of the fate of women and minority judges in Clark County, Nevada in the judicial poll conducted by the *Las Vegas Review-Journal* from 1998 to 2008 raised similar red flags.<sup>21</sup> After controlling for “objective measures of performance quality” such as reversal rates, quality of legal education, extent of judicial experience, ethics history, and initial selection method, the authors concluded that women and minority judges consistently received lower performance ratings than did their white male counterparts.

A more recent, comprehensive examination of this issue by IAALS indicated that these findings are not generalizable to all JPE programs.<sup>22</sup> Taking a straightforward approach, IAALS looked at long-standing official retention evaluation programs in four states—Alaska, Arizona, Colorado, and Utah—to determine whether women judges and judges of color were systematically rated lower than Caucasian male judges. The IAALS study did not directly address the question of whether implicit

20. GARY K. BURGER, *ATTORNEYS' RATINGS OF JUDGES: ST. LOUIS 1998–2006* (2007).

21. Rebecca D. Gill, Sylvia R. Lazos & Mallory M. Waters, *Are Judicial Performance Evaluations Fair to Women and Minorities? A Cautionary Tale from Clark County, Nevada*, 45 *LAW & SOC'Y REV.* 731 (2011).

22. NATALIE KNOWLTON & MALIA REDDICK, INST. FOR THE ADVANCEMENT OF THE AM. LEGAL SYS., *LEVELING THE PLAYING FIELD: GENDER, ETHNICITY, AND JUDICIAL PERFORMANCE EVALUATION* (2012), [http://iaals.du.edu/images/wygwam/documents/publications/IAALS\\_Level\\_the\\_Playing\\_Field\\_FINAL.pdf](http://iaals.du.edu/images/wygwam/documents/publications/IAALS_Level_the_Playing_Field_FINAL.pdf).

biases acted upon respondents' evaluations of judges, but simply whether women and minority judges were at an electoral disadvantage due to phenomena that deserved further exploration. For each state, the data included at least three and as many as five election cycles.<sup>23</sup>

Researchers compared mean scores for judges based on gender and ethnicity on individual survey questions and overall performance criteria. Most of the differences found between evaluation scores for women and men judges, and for minority and Caucasian judges, were small—no more than one-tenth of a point or no greater than 3 percent, depending on the scale of measurement used in that state.<sup>24</sup> However, on a few questions related to legal ability, communication skills, and temperament, women judges or minority judges scored lower than white male judges by two- to three-tenths of a point or as much as 5 percent.

The limited differences found in evaluation scores based on gender and race/ethnicity cannot be definitively attributed to one factor or another, but past research suggests that implicit biases may provide one potential explanation—that those who evaluate judges may unconsciously rely on stereotypes or fixed notions about appropriate roles and behaviors for women and men and for minorities and nonminorities. The few instances where larger evaluative differences were found in the IAALS study were areas where past research suggests implicit biases come into play. Respondents may be invoking these biases in assessing such qualities as judicial competence and judicial demeanor.

The IAALS report makes three research-based recommendations for combating the impact of implicit biases in judicial evaluations. First, JPE programs should be broad-based, particularly in terms of the types of respondents who are surveyed. A significant difference between the evaluation processes featured in the Missouri and Nevada studies and the official JPE programs examined in the IAALS study is that the former consisted solely of attorney surveys, while the latter included nonattorneys, other judges, and jurors as respondents. In fact, the only differences in the IAALS study based on gender and race were found for attorney evaluations of judges; there were no such disparities in evaluations by other respondent groups.

23. These states have reviewed and refined their JPE surveys over time, and individual questions, and overall criteria, have been modified.

24. In Alaska, Arizona, and Colorado, judges are rated on a five-point scale; in Utah, the percentage of "favorable" responses to each question is reported.

The second recommendation, then, is that state court systems and state bar associations redouble their efforts steps to raise awareness of, and promote education about, the potential for implicit biases to impact the judiciary, particularly in this important area. The NCSC offers several excellent resources in this regard.

Finally, it is essential that evaluation surveys be developed in consultation with experts in the field of job-performance evaluation and survey design. In 2010, working with the Illinois Supreme Court Judicial Performance Evaluation Committee, the NCSC underwent a rigorous process of expert consultation, pilot testing, and cognitive interviews to craft attorney and court staff surveys for evaluating the performance of trial judges. IAALS replicated this model in 2013 in creating model surveys for attorneys, trial judges, and court staff to use in evaluating appellate judges.

Two of the most important lessons learned from these survey-development projects regarding negating the potential for implicit biases to influence survey responses were to limit respondents to those who had recent professional contact with the evaluated judge and to employ question wording that reflects readily observable behaviors rather than general attributes or performance criteria. For instance, survey questions such as “The judge’s ruling cited applicable substantive law” and “The judge writes opinions that clearly set forth any rules of law to be used in future cases” are preferable to survey items like “Legal knowledge” or “The judge is competent in the law.” This principle of performance evaluation is another point of departure between the official programs examined in the IAALS study and the Missouri and Nevada studies discussed above.

These best-practice surveys for evaluating trial and appellate judges may be tailored to fit the specific needs and objectives of individual states, and are appropriate for enhancing existing JPE programs or establishing new ones.

## **JPE: LOOKING BACK, LOOKING AHEAD**

In the 15 years since the previous edition of this handbook was published, significant advancements have been made to ensure that JPE programs are as cost efficient, accessible, and effective as possible.

One of the most significant developments has been the expanded use of electronic surveying. Most state-run and bar-sponsored programs

employ electronic surveys for at least some respondent groups. As long as e-mail addresses are available, large respondent populations can be invited to complete JPE surveys, with follow-up as appropriate. This results in meaningful savings on printing and postage. In addition, some state programs now utilize online survey software to tabulate and summarize survey responses, eliminating the expense of hiring an outside contractor. The cost savings from this development are substantial.

Significant changes have also occurred since 2001 in how JPE programs communicate with voters. Perhaps the greatest challenge in using JPE to educate retention election voters is making them aware that the information is available and accessible. This can be the most expensive aspect of a commission's work. A few states have a leg up in this regard, in that evaluation reports are included in state voter guides mailed to each registered voter. The task remains, however, to let voters know the reports are there. In other states, commissions must inform voters of the reports' existence *and* guide them to those reports. Official JPE programs may spend a good portion of their budgets on print and online ads in state and local media to publicize performance evaluation results.

In recent years, the explosion of social media has revolutionized the way that JPE commissions reach out to voters. These commissions are increasingly making use of communication channels such as Facebook, Twitter, and YouTube and the low-cost opportunities they offer to drive voters to JPE websites. Strategic use of social media has the potential to significantly reduce the publicity costs that commissions have faced in the past.

One of the most promising developments in the last decade has been a greater reliance on research-driven performance evaluation processes, particularly when it comes to the design and administration of surveys. In the early days of JPE, bar polls and the pioneer programs in Alaska and New Jersey served as the exemplars for states developing new programs, but recently we have seen some state programs take a step back to consider what can be learned from research in other fields to make JPE as effective as possible. These programs have begun to utilize the work of experts on survey research, social psychology and implicit bias, and job-performance evaluation to ensure the reliability, validity, and fairness of their evaluation processes. National organizations such as the NCSC and IAALS have provided assistance in this regard, developing models and making recommendations for evaluating judges that are based on expert consultation, focus groups, cognitive interviews, and pilot testing.

The JPE chapter in the previous edition of this handbook identified four states as having model programs—Alaska, Arizona, Colorado, and Utah. Over the last 15 years, these states’ programs regularly reassessed all aspects of their evaluation processes to assure that they were as effective as possible, and they made improvements as necessary. In Utah, for example, the JPE commission implemented an innovative courtroom observation program in 2010. The program focuses on procedural fairness, elicits qualitative rather than quantitative feedback from observers, and uses content analysis to prepare reports for each judge. Colorado has worked over the last decade to enhance and refine its process for reviewing appellate judges’ written opinions, increasing the number of decisions to be examined and developing evaluative guidelines and criteria. Arizona’s commission on judicial performance review revisits its public outreach efforts after each election cycle to determine how best to make voters aware of evaluation results for judges standing for retention. These efforts paid off in 2014, when for the first time voters ousted a judge who did not meet performance standards. The Alaska Judicial Council substantially revised its surveys in 2005, switching from paper to mostly web-based surveys.

While there have been a number of developments in the JPE field since 2001, there is one constant: these four states continue to serve as models for other states looking to adopt or improve a JPE program. In addition to providing guidance regarding respondent groups to survey, evaluative tools to employ, and voter outreach opportunities, these states’ programs highlight the imperative of continuous improvement in JPE. JPE programs must, as they have in the past, take advantage of technological advances, lessons learned in other states, and new research and recommendations to maintain evaluation processes that best meet the needs of voters, judges, and the judiciary as a whole.