

Administration & Society

<http://aas.sagepub.com/>

Who Watches the Watchmen?: Evaluating Judicial Performance in the American States

Sharon Paynter and Richard C. Kearney

Administration & Society 2010 41: 923

DOI: 10.1177/0095399709349698

The online version of this article can be found at:

<http://aas.sagepub.com/content/41/8/923>

Published by:



<http://www.sagepublications.com>

Additional services and information for *Administration & Society* can be found at:

Email Alerts: <http://aas.sagepub.com/cgi/alerts>

Subscriptions: <http://aas.sagepub.com/subscriptions>

Reprints: <http://www.sagepub.com/journalsReprints.nav>

Permissions: <http://www.sagepub.com/journalsPermissions.nav>

Citations: <http://aas.sagepub.com/content/41/8/923.refs.html>

>> [Version of Record](#) - Dec 22, 2009

[What is This?](#)

Who Watches the Watchmen? Evaluating Judicial Performance in the American States

Administration & Society
41(8) 923–953
© 2010 SAGE Publications
DOI: 10.1177/0095399709349698
<http://aas.sagepub.com>



Sharon Paynter¹ and Richard C. Kearney²

Abstract

Scholars and human resource practitioners agree that effective performance appraisal systems have clear objectives, reliable and valid appraisal methodology, separation of personal judgments from job-based performance assessments, acceptance by employees, and leadership commitment. Using data from state reports, surveys, case histories, personal interviews, and judicial performance appraisal studies, this article juxtaposes judicial performance evaluation (JPE) and the criteria for effective appraisal systems to address the question of whether judicial independence can be preserved when judges' performance is systematically evaluated by multiple raters. The authors conclude that JPE is an effective performance appraisal tool that can satisfy the need for accountability to the public while protecting judicial independence.

Keywords

judicial performance evaluation, court reform, performance appraisal, performance evaluation

¹East Carolina University, Greenville

²North Carolina State University, Raleigh

Corresponding Author:

Richard C. Kearney, School of Public and International Affairs, North Carolina State University, Raleigh, NC 27695-8102

Email: kearney@chass.ncsu.edu

The Setting: Independence and Accountability in State Judicial Systems

The American judicial system is grounded in two essential tenets: judicial independence and accountability. Judicial independence requires that judges are able to decide cases without bias relying only on the law and facts presented. The Constitution of the United States and, perforce, state constitutions, guarantee the basic rights of individuals including access to, and the impartiality of, the judicial system. Judges should adjudicate freely, without pressure from politicians, bureaucrats, interest groups, or the public. The second, sometimes conflicting, tenet of the U.S. judiciary is accountability to the people. Accountability is manifested in judicial elections and retention votes in merit plan states but is less emphasized in states that select judges through legislative action or gubernatorial appointment. This article considers whether judicial independence can be preserved when judges are subjected to a performance appraisal process intended to offer an avenue for enhancing public accountability. Judicial performance evaluation (JPE) is that tool.

The balance between public accountability and independent decision making is distinctive in the judiciary in comparison to the interplay of these values in the executive and legislative branches. Judges hold the power to check potentially oppressive actions of those in legislative or executive office. They are charged with ensuring preservation of the inalienable rights and freedoms that are hallmarks of American democracy. Sometimes, judges have to make unpopular decisions to uphold the daunting responsibilities of their office. Despite the need for an independent judiciary, courts depend on the other branches of government for resources. State legislatures make most funding allocations for court systems. Moreover, state judges are seated through elections or by appointment of the executive or legislative branches. Given these checks, balances, and influences of the executive and legislative branches, can an independent and accountable judiciary exist? In essence, who should watch the “watchmen” (Sosin, 2000; Winstedt, 1899)? And from a practical perspective, how?

State judges have experienced highly publicized threats to judicial independence (Lubet, 1998). California Chief Justice Rose Bird and two associate justices were removed from the bench in 1986 as the electorate expressed displeasure with a series of controversial rulings expanding rights of convicted felons and overturning 61 death penalty cases. In 1996, supreme court Justice Penny White became the first to lose a retention election in Tennessee. A coalition of groups, including the state Republican Party, targeted White by focusing on a controversial opinion issued on appeal of a lower

court's conviction of a suspect accused of the rape and murder of a 78-year-old woman. Attacked as "soft on crime" by conservatives and the media, she was not retained (see S. Roberts, 2000). In a more recent case, a corporate CEO successfully spent \$2.3 million to unseat a West Virginia Supreme Court incumbent (Hunter, 2004). These examples illustrate the kinds of issues that challenge judicial independence and accountability.

Electorate decisions based on a single issue, opinion, or judgment can seriously affect judicial independence. The ability to interpret the law, manage a caseload, and work effectively within the court system should, ideally, determine whether a judge is qualified to be returned to the bench. However, voters in the aforementioned cases registered their dissatisfaction through the only accountability mechanism available to them, an election. The clash between judicial independence and accountability in California, Tennessee, and West Virginia sent shock waves through other state judiciaries. Even today judges faced with difficult decisions on polarizing issues have been known to observe they "don't want to end up like Penny White."

Similarly, the value of judicial accountability is being seriously debated. Evidence is mounting that campaign donations influence judicial decisions (McCall, 2003; Ware, 1999) and threaten judicial independence. Judicial contests are increasingly politicized, including negative ads, escalating election costs, and other trappings of executive and legislative branch elections (Gleyh, 2003; Maute, 2000; Reid, 1999; Streb, 2007).

An increasingly wide array of stakeholders involved in judicial policy and politics express the view that both independence and accountability are under fire. Displeasure with judges appears to be expanding beyond affronted individuals, opinion groups, and firms to the broader electorate. In November 2006, state referendums in South Dakota, Montana, and Colorado threatened judges with civil and criminal penalties, recall elections, and retroactive term limits, respectively, for case decisions that offended voters. All referendums failed (Kourlis & Singer, 2007), but the very tenets on which the judicial branch is based are being challenged.

Across the United States, judicial races are becoming more politically contentious. Yet voter turnout in these elections typically remains woefully low, no matter whether the contest is partisan, nonpartisan, or retention (Streb, 2007). Certainly, ignorance of judicial candidates' assets, liabilities, and records is partly to blame. The media usually contribute little of substance to judicial elections. One may also assume that a large majority of the state electorate is disinterested in judicial elections (but see Hall, 2006). In this atmosphere, it is not surprising that incumbency is a powerful force except in tightly contested partisan contests or in those instances in which a

controversial ruling on a hot button issue can cost an incumbent his or her seat (Bonneau, 2007). One might reasonably conclude that in the typical context of voter disinterest and nonparticipation, judges are not being held accountable to the people and to those who frequent the courtroom. Where judges fear decisions rendered will lead to interest-group attacks that might threaten their future, independence is also endangered.

From this environment, JPE emerged. This reform is designed to give judges information intended to result in self-improvement and, in some states, also to provide systematic assessments of individual judicial performances to voters as they make decisions about which "watchmen" will sit on the bench. Application of this ancient (Gladden, 1972; Weise & Buckley, 1998) public administration technique to those who wear the robes has diffused across the states for almost three decades. We consider the appropriateness and effectiveness of JPE in the context of judicial reform, and its relevance for the values of judicial independence and accountability.

Judicial Reform

Reform of the courts has tended to occur in response to high levels of public dissatisfaction (Hays & Douglas, 2006). The judiciary, the most traditional and convention-bound of the three branches of American government, has always been insular, self-regulating, and reluctant to reform. Precedent, relative procedural isolation, and organizational fragmentation reinforce resistance to change (Wice, 1995). When reform happened, attorneys, the dominant profession in the judicial system, took the lead and generally controlled the outcome through rules promulgated by the American Bar Association (ABA) or state bar associations (Friedman, 1969; Hays & Douglas, 2006). The salient treatises for contemporary court reforms are written by the ABA, the leading professional organization in the legal field. The ABA's *Standards Relating to Court Organization* (1990) and *ABA Standards for Judicial Conduct* (1990) have served as both touchstones and prime movers for reform.

Although distanced somewhat from the waves of administrative reforms in other public venues (Light, 1997), courts are not immune to the influence of those economic and political forces that shape public administration in the executive branch. The reinvention movement and its legacies, including quality management, benchmarking, productivity programs, and metrics, have affected state judiciaries. Over the past 20 years, increasing attention has been focused on the behavior, decision-making quality, efficiency, case management, and overall performance of judges (Graham, 1993; Stupak,

1991). JPE is an important administrative technique, and court system reform, to hold judges accountable while, in a sense, protecting their independence. It represents a promising portal for public administration to contribute to court reform.

JPE as Performance Appraisal

JPE is in essence a performance appraisal system. Conventionally designed to fulfill at least one of four major purposes, performance appraisals assess skill deficiencies, poor work behavior, or performance problems; determine appropriate pay and classification levels and merit-based pay increases; document behavioral or performance problems for disciplinary actions; and identify needs for employee training or development. All of the above are relevant to judges except salary decisions, which reside in the legislative domain. This is not unimportant: Some of the greatest dissatisfaction with public sector performance appraisal is registered when it is used to make merit pay distributions (Gabris & Mitchell, 1985; Kellough & Selden, 1997). The setting of JPE is challenging and unique; implementation problems are lessened by the salary-setting responsibility of the legislature.

Traits of Effective Performance Appraisal Systems

Despite near universal acceptance and use, effective performance appraisals entail complex techniques that are difficult to implement. A plethora of practical problems, including normal human expectations and interactions (Nalbandian, 1981), generate controversy over their utility and effectiveness (G. E. Roberts & Pavlak, 1996). The problems and challenges impeding successful performance appraisal are well known and extensively documented (e.g., Daley, 1992; Gabris & Ihrke, 2000). People dislike being judged and criticized, and many supervisors are uncomfortable offering personal critiques of subordinates. Nonetheless, scholars (e.g., Daley, 1991) have found that with appropriate design, training, implementation, and leadership, evaluations can be effective.

Human resource management scholars have written prolifically on identifying the traits of effective systems (Brett & Atwater, 2001; Daley, 1985; Dorfman, Stephan, & Loveland, 1986; Levy & Williams, 1998; G. E. Roberts, 1992, 1995). There are five generally agreed on criteria essential to effective performance appraisal systems. Those characteristics are (a) clear objectives, (b) reliability and validity of the appraisal methods, (c) separation of personal judgments and bias from job-based performance assessments,

(d) employee acceptance of the evaluation system, and (e) commitment by organizational leaders to the appraisal process (Cummings & Schwab, 1973; Murphy & Cleveland, 1995). We assess whether a comprehensive JPE process, based on a 360-degree approach, meets the criteria for effective appraisal by examining JPE processes in best-practices states.

360-Degree Performance Appraisal

Multirater, or 360-degree feedback systems, have gained popularity and become a part of the performance appraisal culture in a growing number of organizations across the United States (Waldman, Atwater, & Antonioni, 1998). The potential for employee improvement and development is one reason use of this technique has spread (Brett & Atwater, 2001). Studies show improved performance information leads to capacity building and effective operational strategies, and ultimately to improved organizations (Moynihan & Ingraham, 2004). The emerging consensus is that 360-degree feedback systems can be used to enhance an organization's ability to maximize its human resources (Brett & Atwater, 2001).¹

Effective multirater evaluations solicit information and feedback from individuals who have knowledge of the quality of one or more aspects of an individual's job performance (Daley, 2007). Effectively executed, 360-degree systems are thought to be more balanced, accurate, and useful for employee development than other techniques. Manager and administrator evaluations may use 360-degree systems because these positions are complex, including interactions with other employees at various levels and myriad responsibilities (Brett & Atwater, 2001). In the case of managers and administrators, there are five principal sources of feedback: (a) supervisor, (b) subordinates, (c) peers, (d) the ratee's self-assessment, and (e) clients (e.g., see DeLeon & Ewen, 1997; DeNisi & Kluger, 2000). In some systems, "objective" indicators such as workload and performance measures are used as well.

JPE is generally conducted through surveys of court participants who have direct knowledge of and experience with the judges. These respondents include attorneys, litigants, and for nonappellate judges, jurors and witnesses. Some states also use experts to observe the judge in the courtroom. Experts are usually retired judges or representatives from the JPE commission (L. Cohn, personal communication, April 22, April 23, May 5, 2007; A. J. Conklin, personal communication, August 17, September 26, 2007; J. Howell, personal communication, April 23, May 5, 2007). Judges and commissioners may participate in the process further by meeting with ratees to discuss evaluation results.

Like other appraisal techniques, 360-degree evaluations are susceptible to criticism. They are difficult and expensive to design and implement, consume a significant amount of employee and organizational time, and present problems maintaining confidentiality and anonymity of respondents. Although necessary, training raters is time consuming and expensive. But without properly trained evaluators, the participative process becomes less robust, less effective, and less accepted by employees (DeNisi & Kluger, 2000). Raters sometimes use the opportunity to offer written comments as an avenue to air organizational complaints or personal frustrations rather than to provide specific information to aid in skill development for the employee being rated (Ghorpade, 2000). Moreover, although anonymous ratings made by a broad range of constituent groups are generally perceived to be more accurate and reliable than single-rater systems, these ratings may still lack validity (Ghorpade, 2000). For example, evaluating a person with somewhat opaque managerial responsibilities in a highly specialized job, such as a judge, is difficult, in that it asks diverse raters to remember, retrieve, and integrate complex pieces of information (Ghorpade, 2000).

Although complicated and resource-intensive, multirater feedback systems can play an important role in building a holistic understanding of an individual's performance in the context of organizational effectiveness (Bowman, 1999). Evaluation instruments designed to give raters and ratees specific information from multiple sources can increase the reliability and validity of evaluations and lead to enhanced employee acceptance of the appraisal system. Using multiple sources of information also helps reduce the effects of personal biases—an important consideration when evaluating judges. Because 360-degree evaluations have upward, downward, and horizontal feedback mechanisms, there is also the promise of improved communications and feedback, higher levels of participation, and more meaningful goal-setting (DeLeon & Ewen, 1997; G. E. Roberts, 1995).

JPE is a form of 360-degree evaluation with the foundational elements of an effective appraisal system. JPE programs vary in scope and procedures but most are guided by two overriding goals: judicial improvement and voter education. Through statute, most states using JPE have established a committee or commission to oversee the evaluation process. These groups set the parameters for performance assessment by developing rating tools, identifying raters, determining the scope and depth of evaluations, and providing feedback through personal interviews, written evaluation summaries, and publicly available reports. Because JPE commissions are removed from direct supervision, complex and often negative interpersonal dynamics that sometimes characterize traditional performance appraisal methods are lessened.

Appointment to a judicial evaluation commission or a judicial election process can involve political calculations (Dubosky, 2007). Colorado, for example, does not require equal representation of political parties or any other stakeholder group in the rules governing selection of its judicial performance commission (Colorado Commission, 2008). In Utah, however, the JPE process is managed by a 14-member judicial council including the Chief Justice of the Supreme Court, a Supreme Court Justice, a judge of the Court of Appeals, five District Court judges, two Juvenile Court judges, three Justice Court judges, a state bar representative, and the State Court Administrator, who serves as secretariat to the Council (Utah State Courts, 2008a, 2008b).

When the U.S. Supreme Court decided to allow the same kinds of campaign activities in judicial races as in legislative and executive contests, the door was opened for attack ads, negative campaigning, and special-interest financing to permeate judicial elections (see *Republican Party of Minnesota v. White*, 2002). In 2007 more than \$6 million was spent to win a single Wisconsin supreme court seat and estimates are in the \$20 million range for a similar contest in the 2008 Michigan election ("Reform Judicial Elections This Year," 2008). By incorporating feedback from various sources and stakeholders, JPE ameliorates political considerations in judicial performance assessment.

JPE evaluations could be especially valuable when partisan groups seek to affect voting outcomes in judicial elections. For example, in 1986 judges in Tennessee and Nebraska were unseated after interest groups initiated fierce campaigning that emphasized controversial decisions in single cases (Andersen, 2001). Neither state had a JPE process at the time. Although the election outcome may have remained the same, at least the candidates could have publicized the results of a JPE evaluation based on many case decisions over a period of time.

The opportunity to evaluate a judge irrespective of his or her decisions, in a politically neutral manner, is among the most important advantages of JPE. Neutral criteria are used to analyze how carefully a judge interprets the law, conducts the business of the court as a public administrator, interacts with people in the courtroom, manages the workload, and handles other responsibilities (Institute for the Advancement of the American Legal System [IAALS], 2006). In other words, JPE is intended to measure the full range of judicial performance expectations to gauge individual judge performances against benchmarks as well as other judges (IAALS, 2006). Political considerations can intrude, of course, when raters have an ideological or personal axe to grind. But multiparty participation provides more balanced assessment.

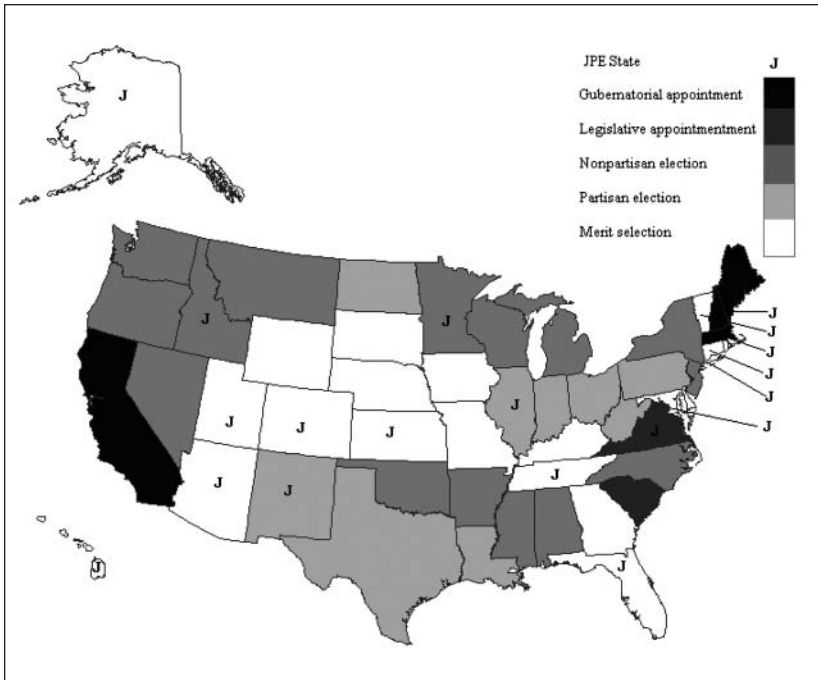


Figure 1. JPE programs and judicial selection systems

In the remainder of this article, we describe our methodology for comparing JPE processes in six leading states to the five criteria for an effective performance appraisal system listed above.

Data and Method

Our research is based on original data gathered through personal and telephone interviews with knowledgeable JPE participants and experts as well as secondary sources including the few existing scholarly studies of JPE and reviews of various state reports by bar associations and research organizations. In 2006 the IAALS at the University of Denver published the first comprehensive report on judicial performance evaluation. Figure 1 relies on IAALS to depict two important variables: (a) the primary vehicle by which judges reach the bench in each state and (b) the 20 JPE states.

The IAALS report identifies six states as “leaders of the comprehensive JPE movement” because each has a formal, comprehensive program featuring

evaluation commissions, multiple raters, anonymous surveys, and public information campaigns. The identified leader states are Alaska, Arizona, Colorado, New Mexico, Utah, and Tennessee. We interviewed experts from these states and used state evaluation tools, voter guides, and other publications to assess JPE as a performance appraisal tool.

We analyze these best-practices states to determine if they use JPE processes that fulfill the essential principles of effective performance appraisal. All six states satisfy the essential criteria for effective performance appraisal systems. However, they do so with different processes and techniques.² They vary according to how evaluation committees are formed, the evaluation process and types of information collected, performance measures, and release of information to the public (IAALS, 2006). Although the sample is small (12% of the state universe and 31.5% of JPE states) and not generalizable to the full population, these six states do represent those identified as engaged in JPE best practices.

360-Degree Evaluation in the Courts: Analysis

Traditionally, tools such as appeal rates, reversal rates, changes in laws and procedure, ethics investigations, reelection processes, and even judicial impeachments have been used to evaluate judicial performance and further accountability in the court system. However, these mechanisms are only available *ex post facto*. None of them facilitates taking proactive steps to improve the performance of individual judges and, by extension, the judiciary as a whole.

Like other multirater assessments, JPE is designed to give individuals comprehensive information about their individual performance (Waldman, Atwater, & Antonioni, 1998). These evaluations operationalize performance by evaluating judges on the basis of rules and ideals set forth in state law and ABA standards. Depending on the level of the court, surveys from jurors, witnesses, attorneys, court staff, professional courtroom participants (law enforcement agents, social workers, peace and probation officers, and others), official courtroom observers, and sometimes other raters are combined to give judges a much more robust sense of personal performance than single-source ratings alone would provide. Ratings and specific comments are made available to the judge being evaluated but individual results are usually not released to the general public. Instead, various judicial commissions use aggregated responses to issue general recommendations to “retain” or “not retain” a judge.

States using partisan or nonpartisan elections to select judges have been reluctant to adopt JPE programs. In part this could be because state-sponsored

evaluations may be perceived by judicial incumbents or challengers as biased, compromising the credibility of the evaluation process. More likely, judges may fear that negative survey results could damage their chances for reelection while leaving their opponents unscathed. Comprehensive JPE adoption is hampered in states using different selection systems for different levels of courts. Oklahoma, for example, appoints judges to the supreme court and the court of criminal appeals through merit selection, but uses non-partisan elections for intermediate appellate court and district court judgeships; municipal judges are appointed by city councils. Intrastate and interstate differences in selection processes make it difficult to create a single set of evaluation techniques that encompasses all the nuances in state court systems. But as the next section makes apparent, JPE has been adopted by states with disparate selection systems.

The Best Practice States

Despite similar objectives, state JPE processes vary widely and by level of court (e.g., higher and lower courts, appellate and general jurisdiction courts). Table 1 shows the components of the JPE programs in the best-practices states. General and special characteristics of each state are presented below before we analyze conformity with essential elements of effective performance appraisal.

Alaska

In 1974 the director of the Alaska Judicial Council (AJC) gained Council support to conduct an information evaluation of state judges. There was no statutory authority for the assessments, no funding for publicizing the results, and though two judges were deemed unacceptable in the rating, voters retained them in the subsequent retention election. Perhaps spurred by that experience and because of efforts of the Council, in 1975 the Alaskan Legislature passed the first state law to create official judicial performance evaluations. In 1976 the initial evaluations were completed (L. Cohn, personal communication, April 23, 2007; L. Cohn, personal communication, September 18, 2008; T. Carns, personal communication, September 18, 2008).

JPE is used in all levels of the Alaskan judiciary. Respondents are questioned on nine categories of the judge's performance using behaviorally anchored Likert-type scale responses. Instruments are tailored to specific respondent groups with eight measurement categories for court staff and six for jurors. For example, jurors are not asked to rate a judge's legal ability because they are not generally trained in the law (AJC, 2008).

Table 1. Common Components of Best-Practices State JPE Programs

State (Year Adopted)	Source of Authorization	Commission Overseeing Evaluations	Commission Membership	Frequency and Release of Evaluation	Participants (Raters)	Method for Information Dissemination
Alaska (1975)	Statutes §22.05.100 §22.07.060 §22.10.150 §22.15.155	Judicial Council	7 members: 3 state bar-appointed, 3 nonlawyers, and CJ Supreme Court; confirmed by legislature	Prior to retention election (results are public)	Attorneys, courtroom observers, court staff, judge (self), jurors, resource personnel	Voter information pamphlet, Judicial Council Web site
Arizona (1992)	AZ Constitution Article 6, §42	Commission on Judicial Performance	30 members: 18 public, 6 lawyers, 6 judges	Midterm (results are confidential) Prior to retention election (results are public)	Attorneys, court staff, judge (self), jurors, litigants (including pro se), resource personnel, witnesses	Voter information pamphlet Commission Web site
Colorado (1988)	§13-5.5-101	Judicial Evaluation Commission (both state and local)	10 members: 4 attorneys, 6 nonattorneys, appointees by CJ Supreme Court, President of Senate and Speaker of House	Midterm (results are confidential) Prior to retention election (results are public)	Attorneys, courtroom observers, court staff, crime victims, judge (self), jurors, litigants, resource personnel, witnesses	Voter information pamphlet, Commission Web site
New Mexico (1997)	NM Supreme Court Order followed by NM constitutional amendment	Judicial Performance Evaluation Commission	15 members: 8 lay people, 7 lawyers. Supreme Court appoints members nominated by legislative, executive, and judicial branches	Midterm (results are confidential) Prior to retention election (results are public)	Attorneys, court staff, crime victims, judge (self), jurors, resource personnel	Voter information pamphlet Commission Web site Local media outlets

(continued)

Table 1. (continued)

State (Year Adopted)	Source of Authorization	Commission Overseeing Evaluations	Commission Membership	Frequency and Release of Evaluation	Participants (Raters)	Method for Information Dissemination
Tennessee (1995)	§17-4-201, §4-29-223	Judicial Evaluation Commission	12 members: 4 state court judges, 2 nonlawyers appointed by Judicial Council, 3 lawyers appointed by Speaker of Senate, 3 appointees by Speaker of House	Prior to retention election (results are public)	Attorneys, court staff, other appellate judges, trial court judges	Commission Web site Local media outlets
Utah (1986)	CJA R3-11, 2-10 6, §78-3-21	Judicial Council with Standing Committee on JPE	14 members: Chief Justice Supreme Court, 12 members elected by judges of various courts, 1 member of Board of Commissioners	Every 2 years	Attorneys, jurors	Voter information pamphlet

Source: Rottman & Strickland (2004).

Nonattitudinal evaluation in Alaska includes calculation of recusal rates, challenge rates, and appellate review of judicial decisions (AJC, 2008). These measures comprise one means of spotting especially egregious problems. For example, if a judge has a high percentage of decisions overturned on appeal, the JPE commission might see this as a red flag worthy of further investigation.

Alaskan judges weigh in on the evaluation process through three qualitative mechanisms: judge-identified respondents (a list the judge furnishes of participants from three trials, three nontrial cases, and any other cases that the judge found significant during his or her most recent term in office), self-evaluations, and voluntary interviews with the JPE commission, called the Judicial Council (AJC, 2008).

Perhaps the most advanced aspect of judicial performance evaluation in Alaska is formal reports made by courtroom observers (L. Cohn, personal communication, April 22, 2007). These evaluators have at least 40 hr of training on judicial evaluation, arrive unannounced in the courtroom, and provide both open-ended written comments and numerical scores (AJC, 2008; L. Cohn, personal communication, May 5, 2007).

Arizona

The selection procedure in Arizona is a bit unusual as judges reach the bench through nonpartisan election in counties where the population does not meet a population threshold of 250,000. Only Pima and Maricopa counties have judges appointed by merit selection and therefore subject to JPE. For courts in these merit selection counties, judicial selection commissions screen candidates for gubernatorial nomination. The governor must appoint a candidate from the nomination list provided (Arizona Commission, 2008).

The instruments begin with an explanation of evaluation criteria. Respondents are then given four choices ranging from *superior* (4 points) to *unacceptable* (no points). Survey questions are designed to rate judges on legal ability, integrity, communication skills, judicial temperament, administrative performance, administrative skills, and settlement activities (Arizona Commission, 2008). These terms can be rather subjective. Consider *judicial temperament*, for instance. The ABA defines temperament as “compassion, decisiveness, open-mindedness, sensitivity, courtesy, patience, freedom from bias and commitment to equal justice” (ABA, 2000, p. 7). The critical need is to clearly link survey questions to specific characteristics of temperament to boost the validity of the instrument.

The Judicial Performance Review Commission uses the combined results to determine whether each judge meets or does not meet judicial standards. To meet the standards, judges must average at least a 2.0 for each category and for each group of respondents, and no more than a quarter of any respondent group can rate the judge as unacceptable or poor in any category (IAALS, 2006). The Commission only reports publicly whether a judge meets the standards; no voter recommendation is offered (Arizona Commission, 2008; A. Corallo, personal communication, May 5, 2007).

Colorado

Coloradan judges have reached the bench through merit selection since passage of a constitutional amendment in 1966. Both appellate and general jurisdiction judges are evaluated. Judicial nominating commissions submit names to the governor for initial appointments. Judges then stand for retention election 2 years later (Colorado Commission, 2008; J. Howell, personal communication, April 23, 2007).

Both local and state commissions govern the Colorado process (J. Howell, personal communication, April 23, May 5, 2007). They are tasked with delivering “fair, responsible, and constructive” evaluations to voters (Colorado Commission, 2008; J. Howell, personal communication, May 5, 2007). Commissioners are appointed by the chief supreme court justice, the governor, president of the senate, and the speaker of the house. As noted, trial and appellate judges are subject to evaluation but recommendations come from different bodies. Appellate judges are evaluated by a statewide commission that ultimately makes recommendations on retention, whereas the trial court judges fall under the purview of 22 local commissions with the same responsibility (Colorado Commission, 2008).

Judges are rated on integrity, knowledge and understanding of the law, communication skills, preparation, attentiveness, control of proceedings, and sentencing practices. In addition to surveys, the commissions review docket and sentencing statistics, conduct personal interviews with judges, make unannounced courtroom observations, and hold public hearings. The results of these multipronged evaluations are assessed and the commissions then deliver a recommendation to voters stated as retain, do not retain, or no opinion. As in Arizona, Coloradan judges receive numerical scores linked to letter grades. Judges must average a score of 2.0—a C average or better—on both attorney and nonattorney surveys for a recommendation of “retain” to be granted (Colorado Commission, 2008; J. Howell, personal communication, April 23, May 5, 2007).

New Mexico

New Mexico only evaluates judges that have been on the bench for at least 2 years after winning a partisan election. Both appellate and district court judges are subject to JPE. Following the initial competitive partisan election, judges are held accountable through retention elections (New Mexico Judicial Performance Evaluation Commission [NMJPEC], 2006).

Halfway through a judge's term and before a retention election, confidential written surveys are distributed to and completed by individuals and professionals who come in contact with judges in New Mexico's courts (IAALS, 2006; NMJPEC, 2008). Using aggregated survey results, the JPE Commission may recommend one of four courses of action: retain, not retain, no opinion, or no recommendation. The latter two options are chosen when there is insufficient information to make a recommendation or when a judge has been in office for a period that is too short for a fair evaluation to be conducted. Retention recommendations are publicized at least 45 days before an election through the JPEC Web site, a widely distributed voter guide; advertisements in print, radio, and television outlets; and other means.

Tennessee

Evaluations in Tennessee currently include only appellate judges. However, stakeholders across the state are beginning to discuss whether JPE should be extended to the trial court level as well (A. J. Conklin, personal communication, August 17, 2007; September 26, 2007; September 28, 2008). One explanation for why state leaders are holding off on adopting comprehensive JPE reviews may be the unusually complex judicial selection procedures.

Tennessee judges are selected differently depending on which court they sit. Nonappellate court judges must run in partisan elections. The governor makes merit-based appointments for appellate judges. Retention decisions are rendered through a yes-no vote in statewide elections 8 years later. This process, known as the "Tennessee Plan," is a modified Missouri plan differing in its application to specific levels of courts. Evaluations are based on six criteria: integrity, knowledge and understanding of the law, ability to communicate, preparation and attentiveness, service to the profession, and effectiveness in working with other judges and court personnel (A. J. Conklin, personal communication, September 26, 2007; Tennessee Judicial Evaluation Commission, 1998, 2000, 2004, 2006).

The JPE commission also uses caseload statistics, judicial self-evaluations, written opinions, and public input but the key part of the analysis is based on survey responses. Survey questions are organized into the headings of oral

argument, written opinions, administrative performance, and general performance. Ratings range from 1 to 10. Performance is described as *poor* with a score of 1 to 2, *fair* from 3 to 4, *adequate* from 5 to 6, *good* from 7 to 8, and *excellent* from 9 to 10. Scores are not released to the public. As a result, the cut-off above which a judge receives a “retention” rating is not publicly available. The evaluation is simply reported as a recommendation for retention, nonretention, no opinion, or no recommendation (A. J. Conklin, personal communication, September 28, 2008; IAALS, 2006; Tennessee Judges Evaluation Report, 1998, 2000, 2004, 2006).

Utah

In Utah, appellate, district, and trial court judges are subject to JPE review. Judges are selected through a merit system and face retention elections after one term in office. Before this election, the Utah Judicial Council conducts an evaluation of the judge’s performance to certify whether he or she should be retained (T. Shea, personal communication, May 5, 2007).

The evaluations measure integrity, knowledge and application of the law, ability to communicate, preparation, attentiveness, control of the courtroom, administrative skills, service to the profession, and effectiveness in working with other court personnel. Utah has a bright line standard for judges to gain a recommendation for retention. For example, a recommendation to retain means a judge got at least 70% favorable responses from nonattorney surveys and 75% from those completed by attorneys. Trial judges must gain favorable responses from at least 75% of jurors. Judges must have no case under advisement for more than 6 months and not more than six cases under advisement for longer than 2 months, and the judge must have completed at least 30 hr of judicial education (*Utah Voter Information Pamphlet*, 2002).

Voter information pamphlets in Utah are much more comprehensive than those released by other states. Publication and dissemination are handled by the lieutenant governor’s office (T. Shea, personal communication, September 19, 2008). Voters receive a complete list of each JPE survey question, a biography for each judge, and a computation of each judge’s score on all attorney survey questions (T. Shea, personal communication, September 19, 2008).

Does JPE Meet the Essential Elements of Effective Performance Evaluation?

JPE structures and processes vary, just as judicial selection systems do. But the critical question is whether JPE is effective, at least in the six leading JPE states for which the most complete data are available. To answer this

question, each of the five major elements of effective performance appraisal is applied to JPE.

Clear Objectives

Understanding what job must be done is a large part of the battle to get it completed. Performance appraisals are no different. Clearly defined goals and objectives give employees and supervisors a means to measure behaviors, task completion, and overall job performance. Goal setting is a critical component of creating performance appraisal systems. Having specific, moderately difficult goals can increase organizational and individual performance as well as employee motivation (G. E. Roberts & Pavlak, 1996).

JPE questions are tied (anchored) to the canons of judicial conduct. Instructions in cover letters accompanying JPE surveys clearly communicate to raters that the purpose of these evaluations is to improve individual judges' performance and, in some states, to inform voters in retention elections. Summary evaluation results are provided to judges reinforcing this dual purpose.

JPE programs benefit from having a professional organization, the ABA, license, train, and guide members of the legal community. The vast majority of judges are attorneys who subscribe to the professional values and rules of practice issued by this body. More specifically, judges are expected to follow the principles stated in the ABA's *Canons of Judicial Conduct*.

The canons are expansive, generalized objectives that can be used to create behaviorally anchored performance appraisal systems with clearly defined goals and objectives. Four are relevant to JPE. Canon 1 states that a "judge should uphold the integrity and independence of the judiciary." Canon 2 is that a "judge should avoid impropriety and the appearance of impropriety in all activities." Canon 3 refers to a judge's responsibilities: "a judge should perform the duties of the office impartially and diligently." The fourth of the seven canons says that "a judge may engage in extra-judicial activities to improve the law and the legal system, and the administration of justice" (ABA Canons of Judicial Conduct; ABA, 1990). The leading JPE states have outlined specific, measurable, and challenging goals and objectives for judges. Figure 2 shows those traits most commonly assessed. Sample questions from the JPE instruments are provided in Table 2.

Validity and Reliability of the Instrument

Validity and reliability are two separate but related concepts. A valid measure is one that measures what it is supposed to measure. The instrument's

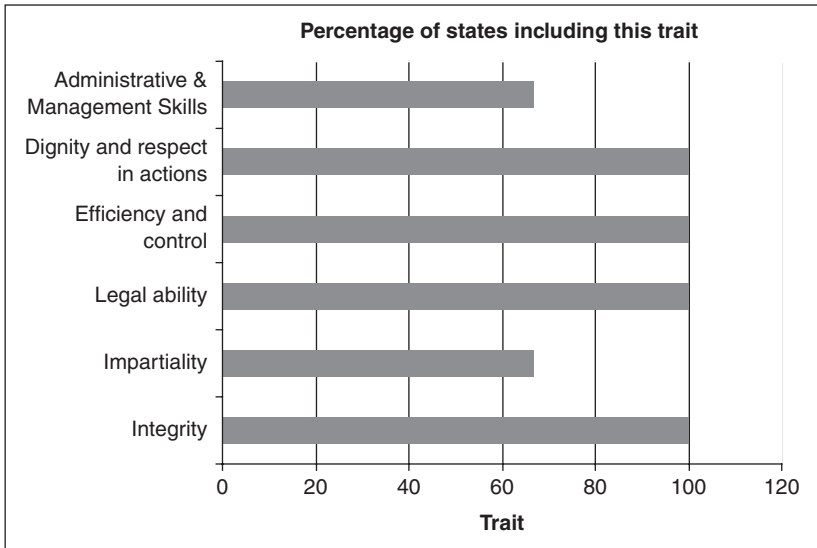


Figure 2. Goals of JPE programs

components should accurately reflect the concept being analyzed—in this case the relevant standards and canons published by the ABA and the state bar association. Reliable measures produce consistent results over time and across respondents. When designing a survey instrument, researchers should ensure that the tool captures what it is intended to measure and that it does so reliably. Consequently, JPE tools should reflect the standards promulgated in the *Canons of Judicial Conduct* or other state or bar association standards and guidelines.

Validity. The first consideration in validity is whether a measure logically represents a particular concept. Certainly, the linkage between JPE measures and the ABA standards establish face validity. Similarly, by explicitly linking standards with measures and providing definitions and behaviorally anchored descriptions of key concepts such as “fairness” and “pays attention,” the JPE instruments appear to have content validity. Statistical indicators such as tracking a judge’s outcomes in managing court dockets and the number of cases overturned on appeal provide evidence of criterion validity.

It is assumed that attorney respondents understand standards set forth in the *Canons* and state laws governing judicial behaviors. Using language similar to that found in the *Canons* and statutes gives attorneys a shared understanding

Table 2. Sample JPE questions from each best-practices state

Canon 1: A judge should uphold the integrity and independence of the judiciary.

- Makes decisions without prejudging the case (AK)
- Makes decisions without regard to possible public criticism (AK)
- Avoided prejudging outcome of case (AZ)
- Dignified (AZ)
- Basing decisions on evidence and arguments (CO)
- Giving reasons for rulings (CO)
- Willing to make decisions without regard to possible criticism (CO)
- Makes decisions without regard to the popularity of the decision (NM)
- Demonstrates courtesy, respect, and collegiality in working with judges and other court personnel (TN)
- Was prepared for the proceedings (TN)
 - Demonstrates knowledge of the substantive law (UT)

Canon 2: A judge should avoid impropriety and the appearance of impropriety in all activities.

- Did the judge insult or undermine either side? (AK)
- Did the judge allow each side equal opportunity to present their case? (AK)
- Did the judge unfairly overrule one side continually in deference to the other? (AK)
- Gives reasons for rulings (CO)
- Conducts self in a manner free from impropriety or the appearance of impropriety (NM)
- Makes decisions without regard to the popularity of the decision (NM)
- Displays sense of fairness, impartiality, and justice (NM)
- Was prepared for the proceedings (TN)
 - Professional behavior is free from impropriety or the appearance of impropriety (UT)

Canon 3: A judge should perform the duties of the office impartially and diligently.

- Did the judge appear to favor either side? (AK)
- Did the judge remain impartial in decisions, statements, attitude, and actions? (AK)
- Clear and logical oral communications/directions (AZ)
- Explained proceedings to the jury (AZ)
- Punctuality in conducting proceedings (AZ)
- Willing to reconsider error in fact or law (CO)
- Having knowledge of rules of procedure and evidence (CO)
- Treats all participants equally (NM)
- Displays sense of fairness, impartiality, and justice (NM)
- Was effective giving information or directions (TN)
- Exhibited courtesy to all persons present during the proceedings (TN)
 - Behavior is free from bias (UT)
 - Demonstrates an ability to perceive legal and factual issues (UT)

Canon 4: A judge may engage in extra-judicial activities to improve the law and the legal system, and the administration of justice. (there were no direct behavioral anchors for this Canon)

Source: State survey instruments.

Table 3. Results of JPE Recommendations and Voter Decisions

	Years of Evaluation	Do-Not-Retain Recommendation	Voter Action
Alaska	1976-2006	9	5 retained 4 not retained
Arizona	1994-2006	1	1 retained
Colorado	1998-2006	14	8 retained 6 not retained
New Mexico	2000-2006 ^a	4	1 retained 3 not retained
Tennessee	1998-2006	0	All retained
Utah	1988-2006	0	All retained

a. New Mexico began a pilot study in 1999.

of survey questions. Jurors and other non-attorney respondents are less familiar with such language, but clear, operational definitions of key terms and concepts helps them assess judges using common constructs.

Predictive validity requires that measures be predictive of future outcomes. Low scores on the JPE instruments would, for example, tend to result in negative assessments and judges being denied a return to the bench when they face an election. The available evidence provides some indications of predictive validity for JPE (Table 3).

Since 1982 every judge who received a retention recommendation has been retained in Alaska; in 2006, one judge failed to gain this recommendation and voters turned him out of office. Arizona voters have received recommendations not to retain two judges since 1992 and both of them also failed to garner enough votes to be reelected. Six Colorado judges receiving do-not-retain recommendations were subsequently rejected by the voters. Other evidence of predictive validity is less direct. After considering negative evaluation results some judges have decided not to stand for retention or reelection. Colorado cites dozens of such instances (IAALS, 2006). Of course, congruence does not always exist between JPE results and election outcomes. In 2004, New Mexico voters chose to retain a judge despite an unfavorable evaluation. A complicating factor in all states is that the overwhelming majority of JPE recommendations are positive. For example, according to Aaron Conklin, assistant general counsel for the Tennessee Administrative Office of the Courts, all the judges standing for retention in Tennessee have received recommendations for retention; in all but one case, the voters agreed with the Judicial Commission's recommendation (A. J. Conklin, personal communication, August 17, 2007; September 26, 2007).

Table 4. Judges' Perceptions of JPE (Colorado Only)

JPE Perception Indicator	Appellate (%)	Trial (%)
Beneficial to professional development	50	85
Detrimental to professional development	21	6
JPE has no effect on judicial independence	41	44
JPE increases judicial independence	24	29
Case management data should be considered in evaluating job performance	72	73
Courtroom observations are <i>somewhat</i> or very useful	N/A	88
Number of survey respondents is major problem (too low)	64	53

Source: IAALS (2008).

Note: JPE = judicial performance evaluation; IAALS = Institute for the Advancement of the American Legal System.

Beyond anecdotal information that judges use performance appraisal results for self-improvement (e.g., Kourlis & Singer, 2007), there is also some empirical evidence. Judges evaluated in a 2004 Washington JPE pilot project reported that the process was “useful” and offered an opportunity for key stakeholder groups to “vent and give feedback on the system” (Brody, 2004). Judges also recommended changes to the process, including “more specific information about negative feedback, more opportunity for specific written feedback” to explain modal responses (Brody, 2004). Coloradan judges have been asked for feedback on JPE. All sitting appellate, district, and county court judges in Colorado were surveyed in early 2008 to get their perceptions of the state’s JPE program. A large majority of respondents consider JPE results to be useful and beneficial to professional development (see Table 4).

Reliability. Reliable performance evaluation systems yield similar results over time and across different raters. JPE surveys use simple, clear instructions and define terms for all respondents, reducing the probability of rater bias and inconsistency. Multiple ratings by a variety of people interacting with the judge also helps assure the reliability of JPE. Alaska is a good case in point. There the Judicial Council provides respondents with precise definitions of the evaluation criteria. These definitions include prompts for objective performance ratings. For example, although questions related to whether the judge paid attention throughout the proceedings could lead to highly subjective responses, the survey instructs respondents to evaluate the judge on task-based traits like whether he or she watched for restless or napping jurors or if the judge paid attention to witnesses, clerks, or others in the courtroom.

Respondents asked to evaluate a judge's case management skills in Colorado are presented with statements operationalizing the concept. Using a 5-point Likert-type scale, respondents rate judges on promptness of rulings, beginning court on time, keeping cases moving, and setting reasonable schedules for cases. Although some of these items contain subjective words such as *reasonable*, having multiple respondents and measures for assessing case management makes this evaluation more valid and reliable.

Separation of Personal Judgments and Bias

As noted above, the various sources and manifestations of rater error are widely recognized. By using 360-degree appraisal methodology and providing definitions for each of the major standards-based assessment criteria, JPE helps raters to avoid such errors and also assists JPE commissions in identifying such errors. The leading JPE states vary in the exact respondent categories, but all six use at least four sets of respondents. New Mexico has the widest scope of respondents of the six states reviewed, surveying law clerks, law professors (for appellate judges), court interpreters, and psychologists in addition to court staff, resource staff, jurors, lawyers, and the judge himself.

Proper rater training is also important. States vary on the level of training provided to raters, but all six offer some. Alaska, New Mexico, and Utah use retired judge evaluators, specially schooled in evaluation procedures, as courtroom observers. Most states rely on written materials such as cover letters, memos, or other publicity to explain the process to respondents.

JPE instruments are designed to measure the presence or absence of traits found in "good" judges. These qualities include legal knowledge, patience, fairness, clarity, and efficiency (IAALS, 2006). Using multiple raters and sources of information gives the JPE commission and the voters the information necessary to understand whether the judges on the benches are indeed good. The process is bolstered by reviewing objective measures of judicial performance. Table 5 shows the types of measures used by the leading states.

Employee Acceptance of the Process

Gaining employee acceptance is a critical component in developing a successful evaluation process. The ratee must believe that the feedback being given is accurate if the results are to be taken to heart and acted on. Ratee perception is shaped in part by the sensitivity and accuracy with which feedback is delivered (Keeping & Levy, 2000; Levy & Williams, 2004; Stone, Gueutal, & McIntosh, 1984). The experience, expertise, and credibility of

Table 5. Types of Objective Measures Used by Leading JPE States

	Recusal Rate	Challenge Rate	Number of Appellate Reviews	Case Management Statistics ^a	Written Opinions
Alaska	X	X	X	X	
Arizona				X	
Colorado		X		X	X
New Mexico				X	
Tennessee					X
Utah				X	X

Source: IAALS (2006).

Note: JPE = judicial performance evaluation; IAALS = Institute for the Advancement of the American Legal System.a.

The case management statistics collected differ from state to state but generally include number of bench trials and trial days, number of jury trials and trial days, sentence modifications, open case reports, and case aging reports.

raters also affect the recipient's willingness to accept the evaluation process. When an evaluator has high levels of these qualities, positive or negative feedback tends to be perceived as accurate and helpful (Stone et al., 1984).

Evaluation can be uncomfortable for anyone. Performance assessment of public officials such as judges, who are both open to public scrutiny and dependent on voters for reappointment to their positions, can be particularly stressful. To help alleviate the stress, each of the six comprehensive JPE states holds a private evaluation meeting between the judge and one or more representatives from the JPE commission, the state judiciary, or the ranks of retired judges. Presiding judges or judges from higher tier courts participate in some states. Communicating evaluation results through knowledgeable and respected peers or supervisors helps evaluated judges accept the process and its results as legitimate. The JPE commissions and the variety of raters would appear to provide highly credible evaluations except, perhaps, to the most resistant judges.

Leadership's Commitment to the Evaluation System

A performance appraisal system is bolstered when an organization's top administrators demonstrate commitment to it (Daley, 1991). When supervisors and managers create an accurate, fair evaluation system designed to improve an employee's performance, and ensure that such a system is implemented and sustained, employees are more likely to participate in the process, accept it as valid and reliable, and work toward improvement.

Symbolic and substantive commitment of judicial leadership to JPE is embodied in judges' direct participation in assessing and coaching peers and lower court judges, as well as through experiencing evaluation of their own performance on the bench. Other indicators of leadership commitment are found in written and verbal communications throughout JPE. In Utah, a signed memo from the chief justice of the state supreme court accompanies each survey asserting its importance to the judicial branch. Explicit public statements and other communications by court leaders—and particularly by the chief justice of the state supreme court—in Alaska, Arizona, and other states send a clear signal of commitment. Further evidence of leadership commitment to JPE comes from the state bar association; without its support, systematic JPE is unlikely to succeed. Finally, state funding for JPE sends a signal that legislative leadership is committed.

Conclusion

Through juxtaposing characteristics of six state JPE leaders with five criteria of effective performance appraisal, we have identified several important findings and conclusions. Unless one considers retention, other elections and *ex post facto* disciplinary-related actions as official forms of evaluation, no other type of judicial evaluation program exists. For this reason, JPE is an important and innovative tool for state court systems. We find that this systematic, objective assessment approach widely used elsewhere in the public sector can be used effectively to evaluate these highest ranking, most influential employees of the judicial branch while preserving the values of judicial independence and accountability. The principles of 360-degree evaluation can be applied to both appointed and elected judges who sit at the apex of judicial hierarchies in the states. By participating in JPE, judges gain the opportunity to improve their performance, while giving other stakeholders in the judicial process a voice that can enhance both intra- and extraorganizational accountability.

The quality of justice quite literally depends on the performance of judges. Judges serve as political actors, dispensers of justice, and court administrators. They exercise enormous influence on public perceptions of justice and the quality of the legal system. They are the *de facto* managers of the courts. Their involvement in managing and administering the business of the court is not only important but essential for the effective administration of justice. Sometimes, particularly in highly charged political environments in which judges participate in an election, perceptions of objectivity are threatened. JPE gives judges, selection commissions, and voters a means to filter out

potentially subjective, biased information that threatens accountability and independence in the judicial branch. When a judge or any other manager exhibits low levels of performance and productivity, the rest of the organization suffers as the weak performance “trickles down the hierarchy” (Daley, 1991). And so JPE is an important tool because it subjects judges to similar kinds of evaluations as other employees while also holding them accountable to the public, governor, or legislature that selected them in the first place.

An interesting observation is that no other elected officials are being held accountable through a systematic performance appraisal technique. Similarly, we could find no evidence that nonjudicial gubernatorial appointees are so evaluated in the states. This raises the intriguing question of whether such officials as agency heads and legislators could—and should—be the subjects of formal performance appraisal with purposes identical to those of JPE, namely, self-improvement and accountability to the voters.

A second finding is that at least in the six JPE states examined in this analysis, 360-degree evaluation appears to be valid and reliable, with positive implications for the independence and accountability of the judiciary. This is important, as the courts are increasingly having to weather attacks from citizens impatient with languorous and tedious court procedures, and outraged that deep-pocket interests and even wealthy individuals seem to be receiving “justice” by “buying a judge” with campaign contributions. JPE has the potential to help overcome citizen aversion and ignorance toward the courts.

We also conclude that public administration research and practice can contribute significantly to the study of state courts and to the improvement of court practices. There are rich opportunities for further contributions in such public administration subfields as budgeting, information technology, policy analysis, performance management, and measurement techniques, among others.

Finally, this article identifies new research opportunities with respect to JPE itself. A majority of states have not adopted JPE. Future research should focus on the diffusion of JPE throughout the states, taking into account the potential social, legal, and political forces that propel—or discourage—adoption. The slow rate of adoption is both a challenge and an opportunity. Because more than half the states have yet to adopt this policy innovation, JPE can be used in a broader context to understand how initial policy decisions are made, what prods actors to begin thinking about the innovation, how members of the systems relate to one another, and what, if anything, JPE replaces.

Decision makers tend to use information that they trust, that they can rely on. Performance appraisal, as represented by 360-degree methodology in this article, is a means to interject merit considerations into any judicial selection process, thereby furthering accountability to those who personally

experience the court system and to the general public, without sacrificing judicial independence. When judges believe evaluations are credible measures of performance that provide useful information, this information can be used to increase individual and thus organizational effectiveness. An independent and accountable judiciary is not only a hallmark of the American legal system, it is a critical element in the balance of power between the three branches of government. Used well, JPE helps us to watch the watchmen.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interests with respect to the authorship and/or publication of this article.

Funding

The authors received no financial support for the research and/or authorship of this article.

Notes

1. For the purposes of this research, the judiciary serves as the “organization” and judges are the “employees” being appraised.
2. Survey instruments and voter guides on file with author and available on request.

References

- Alaska Judicial Council. (2008). *Selecting and evaluating Alaska's judges: 1984 – 2006*. Anchorage, AK: Author.
- American Bar Association. (1990). *Standards relating to court organization: § 1.27. Judicial Performance Evaluation Programs*, ABA. Chicago: Author.
- American Bar Association. (2000, July). *Standards on state judicial selection: Report of the ABA Standing Committee on judicial independence commission on state judicial selection standards*. Retrieved September 17, 2009, from <http://www.abanet.org/judind/downloads/reformat.pdf>
- Andersen, S. S. (2001). Judicial retention evaluation programs. *Loyola of Los Angeles Law Review*, *34*, 1375-1380.
- Arizona Commission on Judicial Performance Review. (2008). *Arizona Commission on judicial performance review*. Retrieved from <http://www.azjudges.info/home/index.cfm>
- Bonneau, C. W. (2007). Campaign fundraising in state supreme court elections. *Social Science Quarterly*, *88*, 68-85.
- Bowman, J. (1999). Performance appraisal: Verisimilitude trumps veracity. *Public Personnel Management*, *28*, 557-576.
- Brett, J. F., & Atwater, L. E. (2001). 360 degree feedback: Accuracy, reactions, and perceptions of usefulness. *Journal of Applied Psychology*, *86*, 930-942.

- Brody, D. C. (2004). The relationship between judicial performance evaluations and judicial elections. *Judicature*, 87, 168-177.
- Colorado Commission on Judicial Performance. (2008). *Commissions on Judicial Performance, years 1998, 2000, 2004, 2006, 2008*. Retrieved from <http://www.cojudicialperformance.com/>
- Cummings, L. L., & Schwab, D. P. (1973). *Performance in organizations: Determinants and appraisal*. Glenview, IL: Scott, Foresman.
- Daley, D. M. (1985). An examination of the MBO/Performance Standards Approach to employee evaluation: Attitudes towards performance appraisal in Iowa. *Review of Public Personnel Administration*, 6(10), 11-28.
- Daley, D. M. (1991). Performance appraisal in North Carolina municipalities. *Review of Public Personnel Administration*, 11(3), 32-50.
- Daley, D. M. (1992). *Performance appraisal in the public sector*. Westport, CT: Quorum.
- Daley, D. M. (2007). *The trials and tribulations of performance appraisal: Problems and prospects in the 21st century*. Unpublished manuscript.
- DeLeon, L., & Ewen, A. J. (1997). Multi-source performance appraisals: Employee perceptions of fairness. *Review of Public Personnel Administration*, 17, 22-36.
- DeNisi, A. S., & Kluger, A. N. (2000). Feedback effectiveness: Can 360-degree appraisals be improved? *Academy of Management Executive*, 14, 129-139.
- Dorfman, P. W., Stephan, W. G., & Loveland, J. (1986). Performance appraisal behaviors: Supervisor perceptions and subordinate reactions. *Personnel Psychology*, 39, 579-597.
- Dubosky, J. E. (2007). Judicial performance review: A balance between judicial independence and public accountability. *Fordham Urban Law Journal*, 34, 315-342.
- Friedman, L. (1969). Law reform in historical perspective. *St. Louis Law Journal*, 13, 351-389.
- Gabris, G. T., & Ihrke, D. M. (2000). Improving employee acceptance toward performance appraisal and merit pay systems. *Review of Public Personnel Administration*, 20, 41-53.
- Gabris, G. T., & Mitchell, K. (1985). Merit based performance appraisal and productivity: Do employees perceive the connection? *Public Productivity Review*, 9, 311-327.
- Geyh, C.G. (2003). Why judicial elections stink. *Ohio State Law Journal*, 64(43), 1-20.
- Ghorpade, J. (2000). Managing five paradoxes of 360-degree feedback. *Academy of Management Executive*, 14, 140-150.
- Gladden, E. N. (1972). *A history of public administration*. London: Cass.
- Graham, C. (1993). Reshaping the courts: Traditions, management theories, and political realities. In S. Hays & C. Graham (Eds.), *Handbook of court administration and management* (pp. 3-25). New York: Marcel Dekker.

- Hall, M. G. (2006, May). *Mobilizing voters in state supreme court elections: Competition and other contextual forces as democratic incentives*. Paper presented at the 6th Annual State Politics and Policy Conference, Lubbock, Texas.
- Hays, S. W., & Douglas, J. W. (2006). Judicial administration: Modernizing the third branch. In G. J. Miller, J. Rabin, & W. B. Hildreth (Eds.), *Handbook of public administration* (pp. 983-1028). New York: CRC.
- Hunter, K. (2004). Election 2004 transforms state supreme court races. Retrieved from <http://www.stateline.org/live/ViewPage.action?siteNodeId=136&languageId=1&contentId=15831>
- Institute for the Advancement of the American Legal System, University of Denver. (2006). Shared expectations: Judicial accountability in context. Available from www.du.edu/legalinstitute
- Institute for the Advancement of the American Legal System, University of Denver. (2008). The bench speaks on judicial performance evaluation: A survey of Colorado judges. Available from www.du.edu/legalinstitute
- Keeping, L. M., & Levy, P. E. (2000). Performance appraisal reactions: Measurement, modeling, and method bias. *Journal of Applied Psychology, 85*, 708-723.
- Kellough, J. E., & Selden, S. C. (1997). Pay-for-performance systems in state government. *Review of Public Personnel Administration, 17*, 5-21.
- Kourlis, R. L., & Singer, J. M. (2007). Using judicial performance evaluations to promote judicial accountability. *Judicature, 90*, 200-207.
- Levy, P. E., & Williams, J. R. (1998). The role of perceived system knowledge in predicting appraisal reactions, job satisfaction, and organizational commitment. *Journal of Organizational Behavior, 19*, 53-65.
- Levy, P. E., & Williams, J. R. (2004). The social context of performance appraisal: A review and framework for the future. *Journal of Management, 30*, 881-905.
- Light, P. C. (1997). *The tides of reform: Making government work 1945-1995*. New Haven, CT: Yale University Press.
- Lubet, S. (1998). Judicial discipline and judicial independence. *Law and Contemporary Problems, 61*(Summer), 59.
- Maute, J. L. (2000). Selecting justice in state courts: The ballot box or the backroom? *South Texas Law Review, 41*, 1197-1224.
- McCall, M. (2003). The politics of judicial elections: The influence of campaign contributions on voting patterns and judicial behavior. *Politics and Policy, 31*, 314-347.
- Moynihan, D. P., & Ingraham, P. W. (2004). Integrative leadership in the public sector: A model of performance information use. *Administration & Society, 36*, 427-453.
- Murphy, K. R., & Cleveland, J. (1995). *Understanding performance appraisal*. Thousand Oaks, CA: Sage.
- Nalbandian, J. (1981). Performance appraisal: If only people were not involved. *Public Administration Review, 4*, 392-396.

- New Mexico Judicial Performance Evaluation Commission. (2006). *New Mexico Judicial Performance Evaluation Commission*. Retrieved from <http://www.nmjpec.org/>
- Reform judicial elections this year [Editorial]. (2008, February 4). *Star Tribune*. Minneapolis-St. Paul, MN, p. A8
- Reid, T. V. (1999). The politicization of retention elections: Lessons from the defeat of justices Lamphier and White. *Judicature*, 83, 68-77.
- Republican Party of Minnesota v. White*, 536 U.S. 765 (2002).
- Roberts, G. E. (1992). Linkages between performance appraisal system effectiveness and rater and ratee acceptance. *Review of Public Personnel Administration*, 5(12), 19-41.
- Roberts, G. E. (1995). Development performance appraisal in municipal government: An antidote for a deadly disease? *Review of Public Personnel Administration*, 15(7), 17-43.
- Roberts, G. E., & Pavlak, T. (1996). Municipal government personnel professionals and performance appraisal: Is there a consensus on the characteristics of an effective appraisal system? *Public Personnel Management*, 25, 379-408.
- Roberts, S. (2000, Winter). Justice goes awry in Tennessee. *The Masthead*. Retrieved from http://findarticles.com/p/articles/mi_m2737/is_4_52/ai_n25031859/
- Rottman, D. B., & Strickland, S. M. (2004). *State Court Organization 2004*. Washington, DC: Government Printing Office.
- Sosin, J. D. (2000). Ausonius' Juvenal and the Winstedt fragment. *Classical Philology*, 95, 199-2006.
- Stone, D. L., Gueotal, H. G., & McIntosh, B. (1984). The effects of feedback sequence and expertise of the rater on perceived feedback accuracy. *Personnel Psychology*, 37, 487-506.
- Streb, M. J. (2007). *Running for judge: The rising political, financial, and legal stakes of judicial elections*. New York: New York University Press.
- Stupak, R. (1991). Court leadership in transition: Fast forward toward the year 2000. *Justice System Journal*, 15, 617-627.
- Tennessee Judicial Evaluation Commission. (2006). *Tennessee appellate judges evaluation report 2006*. Nashville, TN: Author.
- Tennessee Judicial Evaluation Commission. (2004). *Tennessee appellate judges evaluation report 2004*. Nashville, TN: Author.
- Tennessee Judicial Evaluation Commission. (2000). *Tennessee appellate judges evaluation report 2000*. Nashville, TN: Author.
- Tennessee Judicial Evaluation Commission. (1998). *Tennessee appellate judges evaluation report 1998*. Nashville, TN: Author.
- Utah State Courts. (2008a). Court governance. Retrieved from <http://www.utcourts.gov/knowcts/adm/>

- Utah State Courts. (2008b). Judge selection and evaluation. Retrieved from <http://www.utcourts.gov/knowcts/judsel.htm>
- Utah Voter Information Pamphlet. (2002). *General election November 5, 2002*. Retrieved September 17, 2009, from http://elections.utah.gov/GOV_election_pamphletWEB.pdf.
- Waldman, D. A., Atwater, L. E., & Antonioni, D. (1998). Has 360 degree feedback gone amok? *Academy of Management Executive*, *12*, 86-96.
- Ware, S. J. (1999). Money, politics and judicial decisions: A case study of arbitration law in Alabama. *Journal of Law and Politics*, *15*, 645-687.
- Weise, D. S., & Buckley, M. R. (1998). The evolution of the performance appraisal process. *Journal of Management History*, *4*, 233-249.
- Wice, P. (1995). Court reform and judicial leadership: A theoretical discussion. *Justice System Journal*, *17*, 309-321.
- Winstedt, E. O. (1899). A Bodleian MS of Juvenal. *Classical Review*, *13*, 201-205.

Bios

Sharon Paynter is an Assistant Professor in the Department of Political Science at East Carolina University. Her research focuses on the intersection of public and private actors in state policy, including access to institutions such as the state court system. Her work appears in such journals as *Policy Studies Journal* and the *American Review of Public Administration*. After receiving her doctorate from North Carolina State University, she completed a public policy fellowship at Brown University.

Richard C. Kearney is Professor and Director of the School of Public and International Affairs at North Carolina State University. His research is primarily in public administration, particularly human resource management and labor relations, and state and local politics and policy.