

January 2008

Performance Evaluation Program for the Federal Judiciary

Rebecca Love Kourlis

Jordan M. Singer

Follow this and additional works at: <https://digitalcommons.du.edu/dlr>

Recommended Citation

Rebecca Love Kourlis & Jordan M. Singer, Performance Evaluation Program for the Federal Judiciary, 86 Denv. U. L. Rev. 7 (2008).

This Article is brought to you for free and open access by the Denver Law Review at Digital Commons @ DU. It has been accepted for inclusion in Denver Law Review by an authorized editor of Digital Commons @ DU. For more information, please contact jennifer.cox@du.edu, dig-commons@du.edu.

Performance Evaluation Program for the Federal Judiciary

A PERFORMANCE EVALUATION PROGRAM FOR THE FEDERAL JUDICIARY

REBECCA LOVE KOURLIS[†] & JORDAN M. SINGER^{††}

INTRODUCTION

Federal judges enjoy a degree of freedom from structural political constraints unrivaled by nearly all of their counterparts on the state bench. Lifetime appointments shelter district and circuit judges from the fury of periodic elections or reappointment decisions, allowing them to focus on judging and other official duties rather than fundraising, electioneering, or testing the winds of prevailing electoral sentiment. Even federal magistrate and bankruptcy judges not subject to the guarantees of Article III are generally more insulated from politics than their state colleagues, as their appointments and reappointments remain largely internal matters.

Many commentators have praised Article III's guarantees of life tenure and freedom from salary cuts as essential tools to preserve judicial independence.¹ Far less frequently have the commentators explored the impact of these guarantees on judicial *accountability*. Rather, until relatively recently, the prevalent assumption (dating back to the original Federalist debates) has been that "the perceived need for judicial accountability to counterbalance life tenure, nonreducible salaries, and judicial review, began and ended with the impeachment mechanism."² A

[†] Executive Director, Institute for the Advancement of the American Legal System, University of Denver. The Institute is a national, non-partisan organization dedicated to improving the process and culture of the civil justice system. The Institute provides principled leadership, conducts comprehensive and objective research, and develops innovative and practical solutions—all focused on serving the individuals and organizations who rely on the system to clarify rights and resolve disputes.

^{††} Director of Research, Institute for the Advancement of the American Legal System, University of Denver. The authors wish to thank the many judges, lawyers, and scholars whose comments helped us develop the proposal contained in this article, with special thanks to Steve Ehrlich and Russell Wheeler for their insightful comments on earlier drafts.

1. *E.g.*, Luke Bierman, *Beyond Merit Selection*, 29 *FORDHAM URB. L.J.* 851, 864 (2002) ("Life tenure may be the most important ingredient in assuring federal judicial independence."); *see also, e.g.*, Mark A. Behrens & Cary Silverman, *The Case for Adopting Appointive Judicial Selection Systems for State Court Judges*, 11 *CORNELL J.L. & PUB. POL'Y* 273, 305 (2002); Daniel Klerman, *Nonpromotion and Judicial Independence*, 72 *S. CAL. L. REV.* 455, 455 (1999). *But see* Jonathan Remy Nash, *Prejudging Judges*, 106 *COLUM. L. REV.* 2168, 2181 & n.35 (noting that life tenure and the guarantee of no reduction in salary is "only one such template" for establishing judicial independence).

2. Charles Gardner Geyh & Emily Field van Tassel, *The Independence of the Judicial Branch in the New Republic*, 74 *CHI.-KENT L. REV.* 31, 51 (1998); *see also* James E. Pfander, *Removing Federal Judges*, 74 *U. CHI. L. REV.* 1227, 1231 (2007) (noting that Hamilton himself "appears to have embraced impeachment-and-removal exclusivity as a feature of both the New York state constitution and the proposed federal Constitution . . . [and] did not identify any alternative judicial mode by which judges were to be removed from their offices."). *But see* Saikrishna Prakash

reexamination of that assumption, however, has been sparked in the early twenty-first century both by academic commentators and some in Congress. The last ten years alone have produced a host of creative—sometimes outrageous—alternatives to promote federal judicial accountability through (in most cases) a combination of executive and legislative power and populist sentiment. Some such proposals are effectively substance-neutral, most notably replacing life tenure with fixed, lengthy judicial terms.³ Other proposals, however, are aimed at the substance of judicial decision-making, among them several schemes to strip federal courts of jurisdiction to hear certain types of cases.⁴ Prominent politicians have even occasionally threatened impeachment—or worse—for federal judges as a punishment for decisions they did not find appropriate.⁵ Contributing to the tenor of politically “accountable” judges is a federal judicial appointment process that has become increasingly partisan in the last two decades.⁶

Populist-based accountability for judges is precisely what the Founders feared, and should be avoided.⁷ But this does not mean that judges should be exempt from any form of accountability to the citizens they serve. Rather, judges should remain accountable to the public for the

& Steven D. Smith, *How to Remove a Federal Judge*, 116 YALE L.J. 72, 72 (2006) (arguing that the Constitution permits Congress to “enact necessary and proper legislation permitting the removal of federal judges upon a finding of misbehavior in the ordinary courts of law.”).

3. See generally REFORMING THE COURT: TERM LIMITS FOR SUPREME COURT JUSTICES (Roger C. Cramton & Paul D. Carrington eds., 2006); see also Steven G. Calabresi & James Lindgren, *Term Limits for the Supreme Court: Life Tenure Reconsidered*, 29 HARV. J.L. & PUB. POL’Y 769 (2006); James E. DiTullio & John B. Schochet, *Saving this Honorable Court: A Proposal to Replace Life Tenure on the Supreme Court with Staggered, Nonrenewable Eighteen-Year Terms*, 90 VA. L. REV. 1093 (2004).

4. See, e.g., Helen L. Norton, *Reshaping Federal Jurisdiction: Congress’s Latest Challenge to Judicial Review*, 41 WAKE FOREST L. REV. 1003 (2006); Caprice L. Roberts, *Jurisdiction Stripping in Three Acts—Three String Serenade*, 51 VILL. L. REV. 593 (2006).

5. See, e.g., Mike Allen, *DeLay Apologizes for Comments on Judges*, WASH. POST, Apr. 14, 2005 (page unavailable) (quoting House Majority Leader Tom DeLay’s remarks that “the time will come” for federal judges who refused to restore Terri Schiavo’s feeding tube “to answer for their behavior” and that the federal judiciary was “arrogant, out-of-courts, [and] unaccountable.”); Dana Milbank, *And the Verdict on Justice Kennedy Is: Guilty*, WASH. POST, Apr. 9, 2005, at A3 (noting that several Congressional leaders had called for the impeachment of Justice Anthony Kennedy after he authored an opinion forbidding capital punishment for juveniles); see also Editorial, *Unimpeachable Sources—Impeaching Federal Judge Thornton Henderson*, NAT. REV., Feb. 10, 1997 (suggesting that Judge Henderson should be impeached specifically for his decision enjoining California Proposition 209, which sought to prohibit racial preferences in certain programs).

6. See Steven B. Burbank, *Judicial Independence, Judicial Accountability, and Interbranch Relations*, 95 GEO. L.J. 909, 924-25 (2007) (arguing that “there is ample and persuasive evidence from both Supreme Court and lower federal court appointment experience that presidential pursuit of a policy agenda in making judicial nominations (and the reaction to it by Senators of the opposition party) is the chief cause of the politicization of judicial selection at the federal level.”). See also NANCY SCHERER, SCORING POINTS: POLITICIANS, ACTIVISTS, AND THE FEDERAL COURT APPOINTMENTS PROCESS 1-8 (2005).

7. See THE FEDERALIST NO. 78, at 465 (Alexander Hamilton) (Clinton Rossiter ed., 1961) (“The standard for good behavior for the continuance of office in the judicial magistracy is certainly one of the most valuable of the modern improvements in the practice of government. In a monarchy it is an excellent barrier to the despotism of the prince; in a republic it is a no less excellent barrier to the encroachments and oppressions of the representative body.”).

process of adjudication. Judicial decisions, whatever their substantive impact, should be timely rendered, understandable, and supported by clear legal reasoning. Parties and their attorneys should be treated fairly and politely in the courtroom. And the judge should at all times earn the public trust and reputation that naturally comes with his or her position. These considerations locate accountability in actions that should be expected of any judge in any court, regardless of how the judge ascended to the bench or the length of his or her tenure. Embracing accountability for fair and efficient processes may help stave off irresponsible demands for accountability for decisional outcomes.

Accountability based on process measures is not new. Process-oriented criteria are employed regularly at the state court level to measure judicial performance, promote professional development among judges, and educate the public on the importance of accountability for the judicial process as opposed to the substance of specific decisions. At the federal level, however, judicial performance evaluation (JPE) programs remain an untried and (at least in a comprehensive form) unwelcome resource. This need not be the case. The time is ripe to separate the notions of judicial accountability for process and accountability for outcome, and for the federal judiciary carefully to consider process-oriented accountability through a regular performance evaluation program.

This Article begins with a discussion of the purpose and design of JPE programs, gleaned from more than thirty years of experience at the state level. Part II explores the sporadic history of federal JPE, and explains the historical objections to evaluation of federal judges. Part III proposes a series of pilot studies to test different methods of implementing JPE programs. Finally, Part IV discusses some of the more challenging issues presented by the establishment of a federal JPE program, and offers topics for further reflection and research.

I. THE PURPOSE AND DESIGN OF JPE PROGRAMS

Judicial performance evaluation programs are currently in use in various forms in nineteen states, as well as the District of Columbia and Puerto Rico.⁸ The details of these programs vary by jurisdiction, but all are designed to meet three fundamental objectives: (1) to provide constructive feedback to sitting judges to inform their professional development; (2) to educate the public on the work of its judges and foster appropriate expectations about the role of the judge; and (3) where applica-

8. See INSTITUTE FOR THE ADVANCEMENT OF THE AMERICAN LEGAL SYSTEM, SHARED EXPECTATIONS: JUDICIAL ACCOUNTABILITY IN CONTEXT app. A (2006) [hereinafter SHARED EXPECTATIONS].

ble, to provide relevant information to decision-makers concerning the retention or reappointment of judges.⁹

While there is no standard JPE program, most state programs share similar characteristics. Judges are evaluated periodically, either at the end of their terms or at another preset interval.¹⁰ Evaluations are typically conducted by an independent, volunteer commission composed of attorneys, judges, and lay citizens.¹¹ In many states, each branch of government appoints a certain number of members to the commission, thereby reducing the risk of one appointing authority packing the commission with his or her selections.¹² Commission members usually serve staggered terms to further limit any potential mischief by any given appointing authority.¹³

The commission must evaluate judges on predetermined criteria related to the *process* of adjudication rather than to substantive outcomes. Most state JPE programs use the five criteria adopted by the American Bar Association in 1985: legal knowledge, integrity and impartiality, communication skills, judicial temperament, and administrative skills.¹⁴ Guided by these criteria, a present-day commission typically collects a wide range of information on each judge, including surveys of those who interact with the judge in the courtroom (always lawyers, and frequently jurors, witnesses, litigants, or court staff as well), case management data, interview data, information gleaned from direct courtroom observation, and review of the clarity of the judge's written work product.¹⁵ The commission reviews the collected information and composes a detailed

9. Judicial performance evaluation originated in the 1970s and 1980s as a method of providing process-oriented information on a judge's performance to voters in judicial retention elections. It was subsequently adopted by a number of jurisdictions in which judges are subject to periodic reappointment by the governor or state legislature, and even in Massachusetts and New Hampshire, where state judges are appointed for life. See MASS. GEN. LAWS. ch. 211, § 26-26B (2005); MASS. SUP. JUD. CT. R. 1:16 (2008); N.H. SUP. CT. R. 56 (2008).

10. In New Hampshire, for example, trial judges are appointed until retirement or age seventy, and are nevertheless evaluated at least once every three years. See N.H. SUP. CT. R. 56(II)(A) (2008).

11. The size of the evaluation commission varies considerably across jurisdictions, from the seven-member Alaska Judicial Council to the thirty-member Arizona Commission on Judicial Performance Review. See Alaska Judicial Council, Membership, <http://www.ajc.state.ak.us> (last visited Oct. 17, 2008); Arizona Commission on Judicial Performance Review, <http://www.azjudges.info/home/> (last visited Oct. 17, 2008).

12. See, e.g., COLO. REV. STAT. § 13-5.5-102(1)(a)(I)(A) (2008) (dividing appointment authority over the ten-member Colorado state performance commission between the Governor, Chief Justice, President of the Senate and Speaker of the House).

13. See, e.g., S.B. 105, 2008 Gen. Sess. (Utah 2008) (establishing for Utah's new evaluation commission that "At the time of appointment, the terms of commission members shall be staggered so that approximately half the commission members' terms expire every two years.").

14. See generally AMERICAN BAR ASSOCIATION SPECIAL COMMITTEE ON EVALUATION OF JUDICIAL PERFORMANCE, GUIDELINES FOR THE EVALUATION OF JUDICIAL PERFORMANCE (1985) [hereinafter 1985 ABA GUIDELINES]. The ABA reaffirmed these criteria in 2005. AMERICAN BAR ASSOCIATION, BLACK LETTER GUIDELINES FOR THE EVALUATION OF JUDICIAL PERFORMANCE (2005), available at http://www.abanet.org/jd/lawyersconf/pdf/jpec_final.pdf.

15. See SHARED EXPECTATIONS, *supra* note 8, at 20-37 (describing data collected in several states).

report that discusses the judge's perceived strengths and weaknesses on the bench. That report is given to the judge and, if appropriate, also provided to the judge's supervisor and those with the power to determine whether the judge remains on the bench.¹⁶ Reports are also typically made available to the public, either in full or summary form.¹⁷

JPE programs have an established track record at the state level. They are sustainable over many years, even at high volume. In Colorado alone, more than one hundred evaluations of trial and appellate judges are typically conducted every two years,¹⁸ and additional interim evaluations have recently been introduced and formally codified as part of the state's JPE statute.¹⁹ JPE also has positive ripple effects: judges have found that JPE provides useful feedback for their professional growth—information that they could not have otherwise received.²⁰ Furthermore, at least one study has shown that the public has greater confidence in the quality of its judges as a result of JPE programs.²¹ JPE also provides critical information for judicial retention or reappointment decisions, diluting the temptation of voters or reappointment authorities to make such decisions on the basis of specific case outcomes.²²

16. The retention/reappointment authority varies from jurisdiction to jurisdiction. In many states, retention of the judge is left directly to the voters, either in a special retention election in which the judge runs uncontested and must pass a straight up-or-down vote, or in a contested election. In other states, the legislature or governor bear the responsibility for reappointing judges. In Hawaii and the District of Columbia, reappointment and retention decisions are conducted by a commission. See D.C. Code § 1-204.33(c) (2008); HAWAI'I CONST. art. VI, § 3; Hawai'i State Judiciary, Judicial Selection Commission, http://www.courts.state.hi.us/page_server/Courts/2E049BDF320E2D71F0456B57B6.html (last visited Oct. 17, 2008).

17. In New Hampshire and Hawaii, collective reports have been made available to the public that review the judiciary as a whole. See JUDICIARY, STATE OF HAWAI'I, JUDICIAL PERFORMANCE PROGRAM 2007 REPORT (2007) (on file with authors), available at <http://www.courts.state.hi.us/attachment/218D1292A4F6A54DE9973AA6FC/JPP2007.PDF>; Letter from John T. Broderick, Jr., Chief Justice, to John Lynch, Governor, New Hampshire (Jul. 6, 2007) (on file with authors), available at <http://www.nh.gov/judiciary/PerEval/2007-07-03%20final%20report.pdf>. But see Pamela A. Maclean, *More States Evaluating Judicial Performance*, NAT'L L.J. Jun. 2, 2008 (explaining that the New Hampshire Supreme Court will move to individualized reports in 2008). In states with retention elections, full reports are frequently made available on the commission's website, and summary reports are provided in voter guides. See Rebecca Love Kourlis & Jordan M. Singer, *Using Judicial Performance Evaluations to Promote Judicial Accountability*, 90 JUDICATURE 200, 204-05 (2007) (describing methods of public dissemination in each state).

18. Any Colorado judge who is eligible for retention is subject to a full evaluation during his or her retention year. Historically, some judges have chosen not to stand for retention after the evaluation has been completed, for reasons both related and unrelated to the evaluation results. Only the evaluation results of those judges who choose to stand for retention are released to the public. Accordingly, the number of judges who are evaluated is always somewhat higher than the number whose evaluations are made publicly available.

19. See COLO. REV. STAT. § 13-5.5-106.3 (2008).

20. See *infra* nn.142-144 and accompanying text.

21. A seminal 1998 study of JPE programs in four states found that significant majorities of voters who received evaluation information agreed that "the official . . . report adds to my confidence in the quality of judicial candidates [seeking retention]." KEVIN M. ESTERLING & KATHLEEN M. SAMPSON, JUDICIAL RETENTION EVALUATION PROGRAMS IN FOUR STATES: A REPORT WITH RECOMMENDATIONS 41 (1998) (omission in original).

22. See *id.* at 39-40.

The demonstrated benefits of JPE at the state level warrant serious consideration of a similar program for the federal courts. It is true that even the most successful state programs cannot be applied directly to the federal judiciary, and that any federal JPE program would need to be designed to address the unique circumstances of the federal courts. But JPE *does* hold considerable promise for the federal system. Indeed, each of the three major goals of state JPE translates meaningfully to the federal level. First, federal judges, no less than state judges, reasonably could benefit from periodic feedback on their performance based on information gleaned from those who interact with them in the courtroom. Although the Constitution intentionally shelters federal judges from public sentiment to a greater extent than do most state systems, legitimate expectations about a judge's ability to communicate clearly, treat parties fairly, and manage cases effectively apply with equal force to federal and state judges. Information derived from JPE programs might assist not only individual judges, but also Chief Judges, court administrators, and those who design and implement judicial education programs, to capitalize on individual and collective strengths, and address individual and collective weaknesses.

Moreover, if widely disseminated to the public, thoughtful evaluations at the federal level might help to educate the citizenry about its judges. The evaluation process holds the power to be a valuable tool for civic education; regardless of the outcome of any specific judicial evaluation, the routine of evaluating all judges for the same process-oriented skills reinforces to the lay citizen the proper expectations of a good judge. Finally, JPE may prove to be an important asset for those determining the reappointment of magistrate judges, bankruptcy judges, and others not subject to Article III's life tenure guarantees. Simply put, JPE provides decision-makers with information that they would otherwise not have at their disposal; given the choice between having or foregoing relevant, high-quality information, responsible decision-makers should choose to have the information every time.

II. FEDERAL JPE IN CONTEXT

A. The Historical Framework for Process-Oriented Accountability

Some of the principles underlying JPE have been present at the federal level for several decades, even though a sustained JPE program has not. As described in this Part, however, efforts to expand these principles to develop a more comprehensive review of judges' process-oriented performance have fallen short.

1. Nibbling at Accountability: Case Management and Misconduct

Both the federal courts and Congress have emphasized process-oriented judicial accountability measures from time to time, usually in the area of case management. The courts themselves took the lead. As

Chief Justice, Earl Warren noted the negative impact of “[i]nterminable and unjustifiable delays in our courts” on substantive rights,²³ and pushed the Judicial Conference of the United States to study, and eventually recommend to Congress, the establishment of the Federal Judicial Center (FJC) as the research arm of the federal courts.²⁴ Warren Burger, too, bluntly acknowledged as Chief Justice that the federal judicial system needed to explore and adopt better management techniques, and that “[m]ore money and more judges alone is not the primary solution.”²⁵ Certain judges at the district court level subsequently became active proponents of case management among their peers.²⁶ And in 1983, the Federal Rules of Civil Procedure were amended to give district judges greater management control over civil cases.²⁷ With those amendments came the increased expectation of judicial involvement in scheduling events, controlling discovery, and promoting settlement.²⁸

Congress, however, was dissatisfied with the way it perceived some judges to be using (or not using) their case management authority. In 1990 it passed the Civil Justice Reform Act (CJRA), which mandated among other things that the Director of the Administrative Office of the Courts prepare a semiannual report, available to the public, disclosing for each judicial officer the number of motions pending more than six months, the number of submitted bench trials pending more than six months, and the number of cases pending more than three years.²⁹ The CJRA thus created a degree of transparency and accountability regarding the performance of federal judges. But the accountability it created was at once too much and too little. Merely publicizing case processing data about judges artificially elevates the importance of that data over other process criteria. Moreover, information on a judge under the CJRA is not available unless the judge fails to meet the statute’s proscribed outer time limits, and what information is available reflects a mere sliver of the

23. Maurice Rosenberg, *Court Congestion: Status, Causes, and Proposed Remedies*, in *THE COURTS, THE PUBLIC, AND THE LAW EXPLOSION* 29, 31 (Harry W. Jones ed., 1965).

24. See Russell Wheeler, *Empirical Research and the Politics of Judicial Administration: Creating the Federal Judicial Center*, 51 *LAW & CONTEMP. PROBS.* 31, 38-39 (1988).

25. Chief Justice Warren E. Burger, Remarks on the State of the Federal Judiciary (Aug. 10, 1970), in *HOWARD JAMES, CRISIS IN THE COURTS* iv (1971). Burger continued this plea throughout his tenure as Chief Justice. See Warren E. Burger, *Introduction to Symposium, Reducing the Costs of Civil Litigation*, 37 *RUTGERS L. REV.* 217 (1985) (rejecting that additional judicial resources would alone resolve the challenges faced by the federal courts, and arguing that “[j]udicial administration needs tireless, articulate workers.”).

26. See, e.g., Robert F. Peckham, *The Federal Judge as a Case Manager: The New Role in Guiding a Case from Filing to Disposition*, 69 *CAL. L. REV.* 770 (1981).

27. The most prominent of the 1983 amendments were those to Rule 11 (mandating the imposition of sanctions for abuses related to the signing of pleadings and motions), Rule 16 (requiring case management conferences), and Rule 26 (giving the judge authority to keep discovery proportional to the magnitude of the case). See *FED. R. CIV. P.* 11, 16, 26 advisory committee’s notes (1983 amend.).

28. *Id.*

29. 28 U.S.C. § 476 (2008). Given the notion of transparency and accountability inherent in the CJRA, it is ironic that the Director’s semiannual reports are not available to the public on the official U.S. Courts website.

overall picture of a judge's performance with respect to case management.³⁰

Congress's other foray into process accountability for judges occurred in 1980, with the passage of the Judicial Conduct and Disability Act.³¹ That Act established a formal procedure for reviewing complaints "alleging that a judge has engaged in conduct prejudicial to the effective and expeditious administration of the business of the courts, or alleging that such judge is unable to discharge all the duties of office by reason of mental or physical disability."³² The Act charged each chief circuit judge with determining if complaints fell within the Act's coverage, and dismissing those that did not.³³ The Act also charged the Judicial Councils with investigating complaints that the chief judge did not dismiss, and authorized the Councils to take a variety of actions, including imposing a range of statutorily specific sanctions.³⁴ The Act cautioned, however, that "[u]nder no circumstances may the judicial council order the removal from office" of an Article III judge.³⁵

Like the CJRA, the Judicial Conduct and Disability Act addresses only very narrow issues of process-oriented accountability: those concerning formal allegations of misconduct by a federal judge or a judge's inability to discharge the duties of the office for health reasons. Most federal judges never seriously come within its purview.³⁶ The judge whose written order is simply not clear, or whose courtroom manner is abrasive, or whose dockets move at a snail's pace, will fly under the radar of the Act as long as no action rising to the level of formal misconduct is alleged. And the judge whose written orders are careful and thoughtful, and whose manner is unfailingly deserving of respect, will similarly avoid acknowledgment.

30. Some statistical information on the performance of an entire court is available to the public outside the auspices of the CJRA. The Federal Court Management Statistics on the U.S. Courts website provide data on, among other things, each district court and circuit court's overall caseload for the previous five years, actions per judgeships, and median times from filing to disposition and filing to trial for district courts. See generally <http://www.uscourts.gov/fcmstat/> (then follow the District Court hyperlink for the year for which data is sought). The Federal Court Management Statistics, however, do not publicly disclose figures for individual judges.

33. Judicial Councils Reform and Judicial Conduct and Disability Act of 1980, P.L. 96-458, 94 Stat. 2035 (1980) (current version at 28 U.S.C. §§ 351-364 (2006)).

32. 28 U.S.C. § 351(a).

33. *Id.* § 352.

34. *Id.* § 354(a)(1)-(2).

35. *Id.* § 354(3)(a).

36. A recent study found that roughly 650 to 800 complaints were filed annually between 2001 and 2005, with nearly half coming from prisoners. See JUDICIAL CONDUCT AND DISABILITY ACT STUDY COMMITTEE, IMPLEMENTATION OF THE JUDICIAL CONDUCT AND DISABILITY ACT OF 1980: A REPORT TO THE CHIEF JUSTICE 22 (2006). Almost all of the complaints were dismissed, 88% of the time because the allegations related directly to the merits of the case or were otherwise frivolous. *Id.* at 6, 28.

2. Limited Efforts at JPE Programs

Although process-oriented accountability has been addressed to some degree in the areas of case management and judicial misconduct, attempts to implement more comprehensive JPE at the federal level have been sporadic and largely unsuccessful. For thirty years, most of the discussion has centered on evaluating those federal administrative law judges (ALJs) who serve pursuant to the Administrative Procedure Act. In many ways, ALJs were a natural starting point for a federal judiciary hesitant to embrace any form of external evaluation. Although they perform certain judicial functions, ALJs are employees of the executive branch, and indeed are one of the few groups of career federal employees that remain statutorily exempt from performance appraisals.³⁷ Accordingly, proposals to develop a JPE program for administrative law judges have circumvented the thornier issue of Article III independence by couching evaluations as promoting consistency among all executive branch employees.³⁸

Beginning in the late 1970s, several studies suggested that performance evaluations were necessary to assure consistency and efficiency in administrative adjudication. In 1978, the General Accounting Office (GAO) recommended that Congress amend the Administrative Procedure Act to assign responsibility for periodic evaluations of ALJ performance, to be conducted by the Civil Service Commission alone or in conjunction with an ad hoc committee of lawyers, Chief ALJs, agency officials, federal judges, and the Administrative Conference of the United States (ACUS).³⁹ In 1978 and again in 1986, ACUS issued its own recommendations for peer review.⁴⁰ These recommendations emphasized the importance of judicial independence, but also noted that “[m]aintaining the administrative law judges’ decisional independence does not preclude the articulation of appropriate productivity norms or efforts to secure adherence to previously enunciated standards and policies underlying the [agency’s] fulfillment of statutory duties.”⁴¹

In 1992, ACUS issued Recommendation No. 92-7, which proposed among other things that the Chief ALJ be permitted to coordinate development of case processing guidelines for ALJs and conduct annual per-

37. See 5 U.S.C. § 4301(2)(D) (2006) (exempting ALJs from the definition of “employee” for the purpose of performance appraisals). See also Jeffrey S. Lubbers, *The Federal Administrative Judiciary: Establishing an Appropriate System of Performance Evaluations for ALJs*, 7 ADMIN. L.J. AM. U. 589, 590 (1993); 5 C.F.R. § 930.211 (2008).

38. See, e.g., Lubbers, *supra* note 37, at 590-93.

39. GENERAL ACCOUNTING OFFICE, ADMINISTRATIVE LAW PROCESS: BETTER MANAGEMENT IS NEEDED v-vi (1978).

40. ADMINISTRATIVE CONFERENCE OF THE UNITED STATES RECOMMENDATION 78-2, PROCEDURES FOR DETERMINING SOCIAL SECURITY DISABILITY CLAIMS 36 (1978); ADMINISTRATIVE CONFERENCE OF THE UNITED STATES RECOMMENDATION 86-7, CASE MANAGEMENT AS A TOOL FOR IMPROVING AGENCY ADJUDICATION 53 (1986).

41. ACUS RECOMMENDATION 78-2, *supra* note 40.

formance reviews.⁴² The recommendation also proposed a non-exclusive list of criteria for ALJ evaluation, including case processing guidelines (i.e., ALJ productivity and step-by-step goals), judicial comportment and demeanor, and “the existence of a clear disregard of, or pattern of nonadherence to, properly articulated and disseminated rules, procedures, precedents and other agency policy.”⁴³ The recommendation spurred considerable consternation and intense debate.⁴⁴ The performance evaluation program was never implemented, and ACUS itself lost Congressional funding in 1995; one study suggests that ALJs angry with the ACUS proposal were a contributing factor to its demise.⁴⁵

A similar effort to create a performance evaluation program for federal immigration judges—who are Department of Justice employees—was announced in August 2006.⁴⁶ The directive from then-Attorney General Alberto Gonzales described the need for a JPE program to detect unusual reversal rates or backlogs, and emphasized that, in the words of a Justice Department spokesman, “performance appraisals will not be used to tell judges whether to grant or deny relief.”⁴⁷ The announcement nevertheless was met with considerable skepticism by some immigration judges, who voiced concern that the proposal would interfere with their duty to administer their duties neutrally and without political pressure.⁴⁸ Commentators, too, split on whether a performance evaluation program for immigration judges could be constructed in a meaningful way.⁴⁹ To date, the program has not been implemented.

There have also been periodic efforts to introduce JPE into the judicial branch. The Seventh Circuit Judicial Council, for example, has used evaluations to screen sitting bankruptcy judges who are applying for re-

42. ADMINISTRATIVE CONFERENCE OF THE UNITED STATES: THE FEDERAL ADMINISTRATIVE JUDICIARY RECOMMENDATION NO. 92-7, 89 (1992).

43. *Id.*

44. *See, e.g.*, Lubbers, *supra* note 37, at 595-96; James P. Timony, *Performance Evaluation of Federal Administrative Law Judges*, 7 ADMIN. L.J. AM. U. 629 (1993); *see also* Ann Marshall Young, *Evaluation of Administrative Law Judges: Premises, Means, and Ends*, 17 J. NAT'L ASS'N ADMIN. L. JUDGES 1, 54-70 (1997) (proposing a different approach to ALJ evaluation).

45. Toni M. Fine, *A Legislative Analysis of the Demise of the Administrative Conference of the United States*, 30 ARIZ. ST. L.J. 19, 59-61, 96-97 (1998).

46. *See* Press Release, U.S. Dep't of Justice, Attorney General Alberto R. Gonzales Outlines Reforms for Immigration Courts and Board of Immigration Appeals, Aug. 9, 2006, available at 2006 WL 2282541.

47. Nina Bernstein, *Immigration Judges Facing Performance Reviews*, N.Y. TIMES, Aug. 10, 2006, available at <http://www.nytimes.com/2006/08/10/washington/10immig.html>.

48. *See id.*

49. *Compare* Stephen H. Legomsky, *Learning to Live with Unequal Justice: Asylum and the Limits to Consistency*, 60 STAN. L. REV. 413, 468, 469 (2007) (describing as an “especially bad idea”) “[p]erformance reviews that take into account and serve as a criterion for retention and promotion”) with Margaret H. Taylor, *Refugee Roulette in an Administrative Law Context: The Déjà Vu of Decisional Disparities in Agency Adjudication*, 60 STAN. L. REV. 475, 499 (2007) (noting that the 1992 ACUS study “identified several criteria for a system of performance evaluation that appropriately protects decisional independence,” including peer review and external oversight).

appointment.⁵⁰ Such evaluations have consisted of surveys sent to a random sample of one hundred attorneys who had at least two cases before a given judge during the two years prior to evaluation.⁵¹ The Eighth Circuit has used a similar program to evaluate magistrate judges and bankruptcy judges in advance of reappointment decisions.⁵² And in 2003, the Federal Judicial Center assisted the Judicial Conference's Bankruptcy Committee in developing guidelines and surveys for evaluation of bankruptcy judges, for the limited purpose of professional self-improvement.⁵³

To date, however, there have been only two notable efforts to extend JPE to federal district judges. The first was a voluntary program developed in the Ninth Circuit in the early 1980s. The program emerged in response to informal polls conducted by newspapers and bar associations within the Circuit to evaluate federal judicial performance; in the words of the Executive Committee of the Judicial Conference of the Ninth Circuit, a more comprehensive approach under the leadership of the Judicial Conference would "contribute usefully to an effort to make the evaluation of judges as constructive as possible and to avoid the dangers of ill-conceived and sensational 'polls' which merely serve to influence passions."⁵⁴

As this language suggests, the Ninth Circuit project appears to have been initiated and conducted from a strongly defensive posture. In authorizing an Ad Hoc Committee to Study the Evaluation of Federal Judges (the Ad Hoc Committee), the Ninth Circuit Judicial Council simultaneously authorized a parallel Committee to Study the Evaluation of Lawyers.⁵⁵ Furthermore, the Ad Hoc Committee was not authorized to actually evaluate judges, but merely "to evaluate the evaluation of judges."⁵⁶

The Ad Hoc Committee modeled its program on two previous programs in California. The first program used a bar committee to collect attorney complaints about judges. The committee had no power to act on the complaints, but rather passed the complaints along to a committee of judges, who would forward them to the judge in question.⁵⁷ This pro-

50. See DARLENE R. DAVIS, JUDICIAL EVALUATION PILOT PROJECT OF THE JUDICIAL CONFERENCE COMMITTEE ON THE JUDICIAL BRANCH 2 (1991).

51. *Id.*

52. *See id.*

53. FEDERAL JUDICIAL CENTER, 2003 ANNUAL REPORT 11; *see also* Surveys on Behalf of the United States Bankruptcy Court for the Northern District of Illinois (on file with authors), available at <http://www.ilnb.uscourts.gov/Announcements/FJCSurvey.pdf>.

54. Hon. James R. Browning, *Evaluating Judicial Performance and Other Matters*, 90 F.R.D. 197, 198 (1981) (quoting mandate of Ninth Circuit's Ad Hoc Committee to Study the Evaluation of Federal Judges).

55. *See id.*

56. *Id.*

57. *See id.* at 199.

gram was severely limited: it had no transparency, no serious mechanism for accountability other than the notion that “[p]eer pressure . . . would be a far more effective means of correcting judicial deficiencies,”⁵⁸ and no comprehensive scope. The second program considered by the Ad Hoc Committee featured a questionnaire sent to attorneys who had appeared before a district judge in the Northern District of California.⁵⁹ Here again the lack of transparency was trumpeted as a virtue: “The advantages are clear. . . . The judge alone receives the responses. There is no automatic public exposure to put the judge on the defensive and inhibit self-improvement.”⁶⁰

The Ninth Circuit Judicial Council ultimately adopted a voluntary, confidential self-evaluation program for district judges in 1981.⁶¹ Few judges participated. In fact, a 1985 Judicial Council survey found that only nineteen of the 234 judges eligible for the program—less than 8%—had actually undertaken self-evaluation.⁶²

The second effort to evaluate federal district judges came in the form of a pilot program completed in the Central District of Illinois in 1991, under the auspices of the Judicial Conference Committee of the Judicial Branch.⁶³ That district was selected in part because its district judges unanimously expressed interest in the pilot. Indeed, interest was so widespread throughout the district that the pilot program was expanded to include magistrate judges and bankruptcy judges as well.⁶⁴

The pilot was limited in two key respects. First, the only source of evaluation information came from surveys sent to attorneys.⁶⁵ The clerk of the court reviewed a pool of attorneys who had appeared in civil and criminal cases during the eighteen months prior to the study, and sent surveys to a sample of 150 selected attorneys who had appeared before each subject judge.⁶⁶ Jurors, witnesses, and parties were specifically excluded from the study.⁶⁷ Second, the results were entirely confidential and each completed survey was returned directly to the subject judge.⁶⁸ The judges later estimated that the return rate on surveys was about fifty percent.⁶⁹

Despite (or perhaps because of) these limitations, the judges who participated in the pilot project deemed it beneficial. One judge re-

58. *Id.*

59. *See id.* at 199-200.

60. *Id.* at 200.

61. *See* DAVIS, *supra* note 50, at 3.

62. *Id.* at 3-4.

63. *Id.* at 1.

64. *See id.* at 2.

65. *See id.* at 4.

66. *Id.* at 5.

67. *Id.* at 4.

68. *Id.*

69. *Id.* at 8.

marked, "The responses from the bar are an excellent barometer of how we are perceived to be performing our duties."⁷⁰ Another judge stated that the results of the survey are "helpful because they are about as objective an evaluation as we can hope to get."⁷¹

In the final analysis, however, the 1991 pilot study was at best a mixed success. It demonstrated that JPE programs may benefit judges' professional development by providing valuable information about each judge's performance—information that the judge is unable or unlikely to receive in any other format. At the same time, the pilot program clearly failed on two fronts. Most obviously, despite positive reviews from the participating judges,⁷² the program was neither repeated in the Central District of Illinois nor attempted in other jurisdictions. Any momentum toward the design of a more widespread evaluation program was therefore lost. In addition, even if the program had been repeated or expanded in its original form, its extremely constricted scope rendered it of virtually no benefit to judicial training programs, court administrators, or the public. Short of a formal report issued by the Federal Judicial Center the following year,⁷³ no information was disseminated about the results of the program. Accordingly, the public neither learned about the performance of its individual judges nor was afforded the opportunity to see its judges collectively as dedicated public servants striving for continuous professional improvement. Even within the court, where collated survey results might have helped develop new judicial education initiatives or helped the Clerk's Office to anticipate case management issues, no such information was forthcoming.⁷⁴

B. Conceptual Objections to Federal JPE

There are likely many reasons why JPE has not yet succeeded at the federal level, but one key explanation may be anti-evaluation sentiment within the courts themselves. Both conceptual and practical objections have been offered by the courts. We discuss these objections below.

1. Decisional Independence

The most vocal objections to JPE focus on perceived abuses and threats to the judiciary as an institutional actor. The most commonly voiced objection is that JPE, by its very nature, constitutes an assault on a judge's decisional independence.⁷⁵ James Timony, an administrative law judge writing to critique the 1992 ACUS proposal, argued that the

70. *Id.*

71. *Id.*

72. *Id.*

73. *See generally id.*

74. *See generally id.* at 4 ("[T]he subcommittee resolved that the results would remain strictly confidential.").

75. *See, e.g.,* Jacqueline R. Griffin, *Judging the Judges*, 21 LITIGATION 5 (1995).

JPE portion of the report “should be rejected, because it would diminish the decisional independence of federal ALJs and would decrease public acceptance of their decisions.”⁷⁶ Similarly, Denise Noonan Slavin, then-President of the National Association of Immigration Judges, stated that performance review of federal immigration judges was “unwelcome” because it could lead to the public perception that rulings are based on quotas rather than dispassionate application of the law.⁷⁷ And Tennessee Administrative Law Judge Ann Marshall Young has argued that “unless sufficient attention is paid . . . to the need to protect judicial independence on a practical and human basis, the costs of such oversight and evaluation may outweigh any potential benefits.”⁷⁸

Two studies have attempted to measure the perceived impact of JPE on a judge’s independence. A 2008 survey of Colorado judges conducted by the Institute for the Advancement of the American Legal System and Professor David Brody of Washington State University (the “Colorado judges survey”) revealed an almost perfect bell curve of judicial opinion, with 28% of trial judges stating that the state’s JPE program decreases their judicial independence, 44% indicating no effect, and 29% stating that JPE in fact *increases* their independence.⁷⁹ Another study of judges in several states with JPE programs posed the question somewhat differently, asking judges whether they agreed that “[t]he evaluation process undermines my independence as a judge.”⁸⁰ In that study, only 14.5% of judges in Colorado, 22% of judges in Alaska, and 33% of judges in Arizona indicated their belief that their decisional independence was undermined by their state’s JPE program.⁸¹

The survey findings provide reason for both optimism and concern. On the one hand, they demonstrate that a substantial majority of judges surveyed feel that JPE programs do not detract from (and indeed, may increase) their decisional independence. On the other hand, the minority of judges expressing concern about the impact of JPE on their independence cannot be disregarded.⁸² These figures suggest, at least to us, that JPE programs should not be rejected for fear of conflict with decisional independence, but instead should be developed thoughtfully and with judicial input in order to minimize the risk of encroachment on the exercise of independent judgment.

76. Timony, *supra* note 44, at 657.

77. Alfonso Chardy, *Immigration Law: Respect Sought for Busy Judges*, MIAMI HERALD, Sept. 25, 2006, at B1.

78. Young, *supra* note 44, at 7-8.

79. INSTITUTE FOR THE ADVANCEMENT OF THE AMERICAN LEGAL SYSTEM, THE BENCH SPEAKS ON JUDICIAL PERFORMANCE EVALUATION: A SURVEY OF COLORADO JUDGES 31 (2008) [hereinafter THE BENCH SPEAKS]. Nearly two thirds of all judges in the state, at both the trial and appellate levels, responded to the anonymous survey. *Id.* at 2.

80. ESTERLING & SAMPSON, *supra* note 21, at 44.

81. *Id.*

82. *See id.*

Judicial independence is a matter not only of the judge's internal thought processes, but also of public perception. In this respect, JPE programs can put instances of independent, albeit unpopular, judicial decisions into context, thus strengthening judges' willingness to make such decisions. The 2008 state evaluations in Colorado provide a concrete example. In October 2007, Judge James Klein, a district court judge in the state's Twentieth Judicial District, issued a controversial ruling granting a claim for adverse possession.⁸³ Those who disagreed with the ruling immediately branded it as a "land grab"; a term picked up in the media.⁸⁴ Soon Judge Klein was known to most of the public as the "land grab" judge, to the extent he was known to the public at all.⁸⁵ His regularly scheduled performance evaluation, however, properly de-emphasized the single case outcome. The district commission reviewing Judge Klein conducted a thorough evaluation of his overall performance, issuing a report that emphasized his strengths, weaknesses, and opportunities for continued development on the bench.⁸⁶ The commission also thoughtfully put the adverse possession case in the context of his overall caseload and performance:

Judge Klein presided over a highly publicized adverse possession case. The Commission notes that this is only one of over one thousand cases handled by Judge Klein over the past three years. The Commission reviewed Judge Klein's rulings in the case. Judge Klein listened to the testimony presented, visited the site twice, and wrote clear and articulate rulings. Without offering any opinion on the merits of the decision, or whether the decision will be upheld by the appellate court, it is the opinion of the Commission that Judge Klein followed appropriate procedures. Disagreement with the result should not be expressed as unhappiness with Judge Klein's performance.⁸⁷

Judge Klein was ultimately retained by the voters in the November 2008 election.⁸⁸ No matter what the final result at the polls might have been, however, the JPE program served its purpose of locating a single case outcome in the broader context of the judge's overall role.

83. See *McLean v. DK Trust*, Case No. 06-CV-982, at 1 (Colo. 20th Jud. Dist. Oct. 17, 2007).

84. See, e.g., Heath Urie, *Judge to Revisit 'Land-Grab' Case*, BOULDER DAILY CAMERA, Apr. 2, 2008, at A1; Editorial, *Legal Land Grab Should be Overturned on Appeal*, DENVER POST, Nov. 20, 2007.

85. See, e.g., Ryan Morgan, *Judge in Land Case up for Retention*, BOULDER DAILY CAMERA, Apr. 3, 2008, at A1.

86. Judge Klein was appointed to the bench in 2005. See Commissions on Judicial Performance, Honorable James C. Klein (2008), <http://www.cojudicialperformance.com/retention.cfm?ret=210> (last visited Oct. 20, 2008).

87. *Id.*

88. Colorado District Judge Election Results, Denver Post, <http://data.denverpost.com/election/results/district-judge/> (last visited Nov. 21, 2008).

Decisional independence is of course central to the role of the judge. As one commentator has put it, independent decision-making should “be viewed less as a power than as an indispensable responsibility of all judges, at all levels”⁸⁹ Judicial performance evaluation may help judges discharge that responsibility fairly and accurately, and can educate the public on the full and proper role of the judiciary.

2. Life Tenure

Even leaving accountability issues aside, others object that JPE is a needless exercise at the federal level because district and circuit judges cannot be removed simply for underperforming. Judge Timony, for example, has dismissed the notion that state JPE programs are useful models for a federal program, arguing that “[t]he evaluations of state judges usually [are used] in retention elections, a process not relevant to federal ALJs who serve for an unlimited term.”⁹⁰ But this is all the more reason to implement performance evaluations. Promises of continued employment are certainly no excuse for failing to perform to one’s very best ability. Baseball players with guaranteed contracts still work on their swings. Best-selling authors have editors. Self-employed businesspeople send out customer satisfaction surveys. The position of federal judges should be no different. They are appointed to the bench based on a proven combination of skill, experience, and future promise; part of fulfilling that promise is a commitment to the public to grow in the job.

Moreover, for federal judges who do not have life tenure, such as magistrate judges and bankruptcy judges, JPE may assist not only with professional development, but also with reappointment decisions. Full-time federal magistrate judges serve eight-year terms; part-time magistrate judges four-year terms.⁹¹ Terms are renewable with the concurrence of the majority of district judges in a district court, or by the chief judge if there is no such concurrence.⁹² Bankruptcy judges similarly serve fixed terms of fourteen years, renewable by the Court of Appeals for the circuit in which they serve.⁹³ JPE is especially suited for these judges, because it could provide critical information about the judge’s performance to the relevant decision makers in advance of a reappointment decision. We discuss one possible application of JPE to reappointment decisions in Part III.

3. Public Perception

A final conceptual objection to JPE relates to the potential release of evaluations into the public domain. The concern is that rather than fo-

89. Young, *supra* note 44, at 27.

90. Timony, *supra* note 44, at 641.

91. 28 U.S.C. § 631(e) (2008).

92. *See id.* § 631(a).

93. *Id.* § 152.

ocusing on the judges who receive excellent reviews, or even the judges who demonstrate marked improvement in one or more areas, some in the public sphere will emphasize the judge who does badly or who receives particularly harsh comments.⁹⁴ Where informal polls of attitudes toward judges constitute the sole basis for judicial evaluation, there is indeed an increased risk that judges will be inappropriately ranked, or that specific evaluation results will be taken out of context.⁹⁵ However, when evaluations are based on a broad set of process-oriented criteria and are grounded in credible information from a wide variety of sources, and the process itself is transparent, the risk of media sensationalism or public overreach has the potential to be greatly reduced.

There is in fact some evidence that when the judiciary publicly supports a JPE program and makes the results broadly available, it gains the respect and confidence of the mass media—and perhaps by extension, the public. After a bar-sponsored program released evaluations of trial court judges in Pierce County, Washington, in June 2008, the county's largest newspaper ran five different stories on the evaluations. None of those articles focused exclusively on judges who did poorly (although they did mention those judges whose overall evaluations were particularly strong or weak),⁹⁶ and several explicitly praised the judiciary for its increased commitment to transparency and public service. As one editorial put it, "Naturally, sitting judges don't much like getting report cards, but Pierce County's judges cooperated admirably with the bar's rating process. The bench wins more respect when it acknowledges that its members should be held accountable for performance."⁹⁷

Moreover, the Colorado judges' survey suggests that judges who have been through a comprehensive JPE process at least once strongly support providing evaluation results to the public. Nearly 69% of trial judges indicated that they have no difficulty with Colorado's current method of disseminating information to the public, which consists of

94. See, e.g., Griffin, *supra* note 75, at 61-62 ("[A]ny [state] judge who is given a 'do not retain' [recommendation] has no access to information on why or how the decision was made, and he is unlikely to have the resources to mount a response . . ."). This statement is incorrect. In most comprehensive JPE states, each judge receives an extensive report compiling all the data on his or her performance before the evaluation is even released to the public. Colorado allows judges who disagree with a recommendation to seek a second interview with the evaluation commission, and, if necessary, to write a short rebuttal statement to go to the voters prior to the evaluation's release. See S.B. 08-54, 66th Gen. Assem., 2d Reg. Sess. (Colo. 2008) (codified as amended at COLO. REV. STAT. § 13-5.5-106(1)(a)(V) & -(2)(a)(V) (2008)). Arizona also enhances transparency by requiring that the ultimate vote on whether the judge had met performance standards be taken publicly. See ARIZ. COMM'N ON JUDICIAL PERFORMANCE REVIEW, RULES OF PROCEDURE FOR JUDICIAL PERFORMANCE REVIEW IN THE STATE OF ARIZONA, 6(c)(3) (2006), available at <http://azjudges.info/about/procedure.cfm>.

95. See Browning, *supra* note 54, at 199 (discussing a San Francisco Bar Association poll in the late 1970s that led to rankings of individual judges in the press).

96. See, e.g., Editorial, *Bar's Judicial Ratings Will Aid the Voters*, NEWS-TRIBUNE (Tacoma, Wash.), June 3, 2008, available at <http://www.thenewstribune.com/opinion/story/379136.html>.

97. *Id.*

posting full evaluations on the state commission website, and providing short summaries of each evaluation in a voter guide.⁹⁸ The judges who provided comments in the survey uniformly indicated that public dissemination, and efforts to educate the public about the JPE process, could in fact be even more extensive.⁹⁹ If judges who face retention elections mere months after their evaluation favor such efforts to publicize the results, federal judges with life tenure should be comfortable with the release of their evaluations as well.

C. Practical Objections to Federal JPE

1. Cost

One frequently raised objection to JPE programs, even among those who support evaluations in principle, is the cost associated with a regular and ongoing JPE program.¹⁰⁰ Surveys must be sent out and responses tabulated. Where commissions are used, members may have to travel and results must be disseminated. Particularly when taxpayer money is at stake, the cost of a new program is never a matter to be taken lightly. However, the cost of a JPE program for a federal district need not be prohibitively expensive. The 1991 pilot program in the Central District of Illinois reported very few costs, and concluded that “the cost of a similar evaluation program in a large district would most likely be minimal.”¹⁰¹ While a full-scale, nationwide federal JPE program would certainly incur something more than “minimal” costs, evidence from existing programs suggests that it could be done in a cost-effective manner.

Detailed cost considerations usually begin with surveys. Some state programs use private polling companies to design and circulate surveys and tabulate responses, which ensures a high level of professional competence in survey methodology. High quality polling can also be achieved, however, through lower cost means. Some state programs complete their polling through local universities,¹⁰² which promises high quality methodology with potentially less expense. Others have developed electronic surveys,¹⁰³ which eliminates mailing and copying costs, and allows for results to be tabulated on a rolling basis. Existing commercial survey software might well be suitable for use at least in federal

98. THE BENCH SPEAKS, *supra* note 79, at 26.

99. *Id.*

100. See generally Griffin, *supra* note 75, at 5-7.

101. DAVIS, *supra* note 50, at 7.

102. For example, Alaska conducts its JPE surveys through the University of Alaska Anchorage, Virginia uses Virginia Commonwealth University, and a recent pilot program in Pierce County, Washington relied on surveys conducted through Washington State University.

103. The Commonwealth of Massachusetts, for example, developed in-house an electronic system to survey attorneys. Respondents complete surveys for individual judges on a secure, encrypted website, and results are automatically aggregated by judge. Mona Hochberg, Judicial Performance Evaluation Coordinator, Mass. Supreme Judicial Court, Remarks at IAALS Symposium: Judicial Performance Evaluation: Strategies for Success (Aug. 6, 2008) (copy of presentation on file with authors).

pilot programs. Costs associated with more robust JPE programs, such as travel and public dissemination of results, can also be reduced or even eliminated with modern communications technology—teleconferencing and videoconferencing can reduce the number of required face-to-face meetings for the commission, evaluation results can be posted on the court's website at minimal cost,¹⁰⁴ and so on. Reasonable options for financing JPE programs can also be explored, perhaps through a modest raise on application fees for admission to practice in a federal district court.

Actually predicting costs is difficult, at least before thorough pilot studies are undertaken. Data on the cost of JPE programs at the state level are instructive but not dispositive. Massachusetts runs a relatively limited program based on electronic surveys, with no evaluation commission or publication of results, for the cost of one full-time employee and some minor overhead costs.¹⁰⁵ Alaska evaluates anywhere from ten to thirty judges each election cycle, spending \$2000-4000 per judge for surveys, travel, materials, and dissemination inclusive;¹⁰⁶ if staff time were to be factored in, the per judge cost would roughly double.¹⁰⁷ Virginia's JPE program currently spends about \$5000 per judge for surveys, but costs are expected to drop in the future to the range of \$3500-4000 per judge.¹⁰⁸ These programs also benefit from economies of scale; the more judges evaluated during a particular cycle, generally the lower the per-judge cost.¹⁰⁹

2. Risk of Politicization

Another common objection to JPE goes like this: "I support the idea of evaluating judges, certainly for purposes of self-improvement, but if we leave the evaluation to those outside the judiciary even the most carefully designed process is bound to inject politics into a system where none should exist."¹¹⁰ Judges alone, the argument goes, can be trusted to understand the roles and responsibilities of the judiciary, and to reach

104. For excellent examples of state court websites describing their JPE programs and recent results, see Alaska Judicial Council, <http://www.ajc.state.ak.us> (last visited Oct. 17, 2008); Arizona Commission on Judicial Performance Review, <http://www.azjudges.info> (last visited Oct. 17, 2008); Colorado Commissions on Judicial Performance, <http://www.cojudicialperformance.com> (last visited Oct. 17, 2008); Kansas Commission on Judicial Performance, <http://www.kansasjudicialperformance.org> (last visited Oct. 17, 2008); Supreme Court of New Mexico Judicial Performance Evaluation Commission, <http://www.nmjpec.org> (last visited Oct. 17, 2008).

105. See Hochberg, *supra* note 103.

106. Larry Cohn, Executive Dir., Alaska Judicial Council, Remarks at IAALS Symposium: Judicial Performance Evaluation: Strategies for Success (Aug. 6, 2008) (notes from presentation on file with authors).

107. *Id.*

108. Edward Macon, Assistant Executive Sec'y and Counsel, Supreme Court of Virginia, Remarks at IAALS Symposium: Judicial Performance Evaluation: Strategies for Success (Aug. 6, 2008) (notes from presentation on file with authors).

109. Cohn, *supra* note 106.

110. See, e.g., Griffin, *supra* note 75, at 7.

conclusions about strengths and weaknesses in an objective and apolitical manner. Underlying this objection is the fear that various governmental appointing authorities will choose commission members who will evaluate judges on the basis of case outcome rather than adjudicative process.¹¹¹

Proponents of this view, however, are unable to cite to actual examples of politicized JPE programs. Instead, they argue by analogy. A recent article by Justice Charles Wells of the Florida Supreme Court, for example, argued that Florida was right to reject a comprehensive JPE program because the state legislature had changed the statutory composition of its judicial *nominating* commission in a way that “increased the potential for political influence in the selection of judges.”¹¹² From this starting point, Justice Wells extrapolated the conclusion that “there can be no bulletproof guarantee that the judicial evaluation body will remain free of legislative or executive influence.”¹¹³

It is certainly true that distrust between the courts and the legislative branch is broad and deep.¹¹⁴ Recent concerns that proposed legislation for an Inspector General for the federal judicial branch might result in additional scrutiny of judges whose decisions are unpopular with Congress, notwithstanding statutory admonitions to the contrary, has not helped assuage this distrust.¹¹⁵ However, independent JPE commissions should be seen as a possible solution to concerns over legislative or executive encroachment—not something to be discarded because of those concerns. In over thirty years of operation, there are no clear-cut examples in the popular or scholarly literature of state JPE commissions evaluating a judge on the basis of anything other than established, process-oriented criteria. We are certainly not aware of any examples in which a commission systematically targeted judges based on a particular ideology or approach. Rather, the outward politicization of judges and judicial decision-making occurs more frequently in jurisdictions that do *not* have judicial performance evaluation programs.¹¹⁶

The list of judges or entire judiciaries that have been targeted for political reasons in jurisdictions lacking JPE programs is lengthy and stretches out over more than two decades. At the state level, California Chief Justice Rose Bird and two of her peers did not retain their seats on

111. *See id.*

112. Charles T. Wells, Editorial, *Viewpoint: The Inherent Danger of Judicial Evaluation Commissions*, JACKSONVILLE DAILY RECORD, Jan. 7, 2008, available at http://www.jaxdailyrecord.com/showstory.php?Story_id=49192.

113. *Id.*

114. For a discussion of a recent project to promote effective communication between Congress and the courts, see Robert A. Katzmann & Russell R. Wheeler, *A Mechanism for Statutory Housekeeping: Appellate Courts Working with Congress*, 9 J. APP. PRAC. & PROCESS 131 (2007).

115. Russell R. Wheeler & Robert A. Katzmann, *A Primer on Interbranch Relations*, 95 GEO. L.J. 1155, 1171 (2007).

116. *See Kourlis & Singer, supra* note 17, at 202.

the state supreme court in 1986 after an extensive and politicized public non-retention campaign. The three justices, who had no evaluations to document their broader judicial performance (the California judiciary had discussed but declined to adopt a JPE program in the early 1980s),¹¹⁷ were left to defend themselves armed only with the esoteric notion of “judicial independence”—which public polling showed “was the one message that would not work.”¹¹⁸ In 1996, Justice David Lanhier of the Nebraska Supreme Court and Justice Penny White of the Tennessee Supreme Court were separately removed from the bench in highly politicized retention elections.¹¹⁹ Once again, neither justice was able to point to objective evaluations from an independent commission to defuse the political rhetoric. Justice Lanhier limited his active campaigning to a bare minimum in order “to maintain the dignity of the office,” a strategy that failed.¹²⁰ Justice White adopted Rose Bird’s strategy of emphasizing the importance of judicial independence, which unfortunately produced the same result.¹²¹ Justice White has since become an articulate supporter of JPE programs as a bulwark against politicization and for judicial independence, noting that “[u]ndoubtedly, much of the success of those who seek to destroy judicial independence results from the lack of available information upon which to base one’s decision in judicial elections.”¹²²

To be clear, we are not asserting that JPE alone can inoculate the judiciary against politicization efforts. Local and national political culture would seem to have the most powerful impact on the existence and intensity of political attacks on the judiciary, and JPE by itself cannot change a poisoned cultural dynamic. But if JPE is not a vaccine, it is perhaps at least preventive medicine. Efforts to hold judges “accountable” for particular case outcomes appear more likely to find purchase in jurisdictions where process-oriented accountability measures are not publicly available. The “JAIL 4 Judges” initiative¹²³ emerged in South

117. See REBECCA LOVE KOURLIS & JORDAN M. SINGER, A FRESH LOOK AT JUDICIAL PERFORMANCE EVALUATION IN CALIFORNIA 4-5 (2007), available at <http://www.du.edu/legalinstitute/news/CA%20JPE.html>.

118. John T. Wold & John H. Culver, *The Defeat of the California Justices: The Campaign, The Electorate, and the Issue of Judicial Accountability*, 70 JUDICATURE 348, 350 (1987); see also Bill Zimmerman, *The Campaign that Couldn’t Win: When Rose Bird Ran Her Own Defeat*, L.A. TIMES, Nov. 9, 1986, at V1 (noting that “[t]o base a political campaign on the independence of the judiciary was to commit electoral suicide”).

119. See Traceil V. Reid, *The Politicization of Retention Elections: Lessons from the Defeat of Justices Lanhier and White*, 83 JUDICATURE 68, 76-77 (1999).

120. *Id.* at 72.

121. *See id.*

122. Penny J. White, *Judging Judges: Securing Judicial Independence by Use of Judicial Performance Evaluations*, 29 FORDHAM URB. L.J. 1053, 1076 (2002).

123. The proposed initiative, styled as Constitutional Amendment E, would have allowed a thirteen-member “Special Grand Jury” to expose judges and prosecutors, as well as citizens serving on juries, school boards, county commissions, or in similar decision-making capacities to fines and jail—and strip them of public insurance coverage and up to half their retirement benefits—for making decisions that break rules defined by the special grand jurors. See CHRIS NELSON & KEA

Dakota, a state without a JPE program, in 2006. When that effort proved unsuccessful,¹²⁴ proponents proceeded with efforts to get the same measure on the ballot in Florida (a state that recently rejected any public dissemination of its JPE results)¹²⁵ in 2008.¹²⁶ Increased attacks on the state judiciary have also occurred recently in Missouri,¹²⁷ which similarly has lacked a formal JPE program.¹²⁸ Once again, we do not wish to suggest that the mere absence of a formal JPE program caused these attacks; rather, it is sufficient to observe that more existential threats to the independence of the judiciary tend to arise in states where public accountability through a JPE process is wanting.

By contrast, judges in JPE jurisdictions tend to be less subject to specialized attacks. When these attacks do occur, evaluation results—or simply the existence of a JPE program—have been used to emphasize the judges' adjudicative skills and depoliticize special interest messages. In 2006, for example, a ballot initiative was introduced in Colorado which sought to term-limit all of the state's appellate judges.¹²⁹ The initiative was retroactive and would have immediately removed nineteen of the state's twenty-six appellate judges from the bench, regardless of their experience and abilities. A spirited public education campaign, emphasizing among other things the fact that Colorado already had a system for evaluating (and, if necessary, removing) judges in a much more precise fashion, helped to defeat the initiative at the polls.¹³⁰ A follow-up study of Colorado voters found that only 18% of those who voted for judicial

WARNE, SOUTH DAKOTA 2006 BALLOT QUESTIONS (on file with authors), *available at* <http://www.sdsos.gov/elections/voteregistration/electvoterpdfs/2006SouthDakotaBallotQuestion-Pamphlet.pdf>. The proposed amendment was designed to apply retroactively. *Id.*

124. Due in large part to an extensive public campaign, the proposal was ultimately defeated by a 9-1 margin. See SOUTH DAKOTA SECRETARY OF STATE, GENERAL ELECTION OFFICIAL RETURNS FOR BALLOT QUESTIONS (on file with authors), http://www.sdsos.gov/elections/voteregistration/pastelections_electioninfo06_GEBallotquestions.shtm (last visited Oct. 17, 2008).

125. See Letter from Peter D. Webster, Chair, Comm. on Judicial Evaluations, to R. Fred Lewis, Chief Justice, Fla. Sup. Ct. (July 10, 2007) (on file with authors).

126. See Florida Department of State, Division of Elections, Initiative No. 02-06 (2002), *available at* <http://election.dos.state.fl.us/initiatives/initdetail.asp?account=35025&seqnum=1>.

127. As one example, Missouri Governor Matt Blunt used part of his 2008 State of the State Address to encourage the state legislature to “close the door” on courts who have “hijack[ed] the powers to tax and spend,” even though no Missouri state court had raised such an issue in an opinion. Governor Matt Blunt, 2008 State of the State Address (Jan. 15, 2008) (on file with authors), *available at* http://governor.mo.gov/State_of_the_State_2008.pdf. Missouri is now the center of a firestorm concerning the best form of state judicial selection, and legislative threats to discontinue the Missouri Plan—the first state merit selection system implemented in the country—continue.

128. On February 29, 2008, the Supreme Court of Missouri created JPE committees at the trial and appellate levels by court rule, pursuant to its constitutional authority. The basis for the rule was a Report of the Missouri Judicial Evaluation Survey Committee. The program went into effect almost immediately, with the first set of reports and recommendations scheduled to be released in September 2008. See Letter from Dale C. Doerhoff, State Chair, Missouri Judicial Performance Evaluation Comms., to Participants at IAALS Symposium: Judicial Performance Evaluation: Strategies for Success (Aug. 5, 2008) (on file with authors).

129. See STATE OF COLORADO, ANALYSIS OF THE 2006 BALLOT PROPOSALS 7-8 (2006).

130. The proposed initiative gained only 43% support in the November 2006 election. See STATE OF COLORADO, OFFICIAL PUBLICATION OF THE ABSTRACT OF VOTES CAST FOR THE 2005 COORDINATED, 2006 PRIMARY, 2006 GENERAL 140 (2006).

term limits were aware that judges were evaluated by an independent commission; by contrast, 41% of those who voted against term limits knew specifically about the state's JPE commissions.¹³¹ Far from injecting politics into the evaluation process, independent JPE programs—when they are publicly known and understood—have a tendency to serve as a bulwark against political attacks on the judiciary, making such proposals less likely to pass public muster.

III. DEVELOPING A MODEL FOR FEDERAL JPE THROUGH PILOT PROGRAMS

In this Part, we propose a series of different pilot studies to test the benefits of JPE at the federal level. We suggest several alternatives: (1) a pilot program designed to elicit confidential feedback for federal district, magistrate, and bankruptcy judges strictly to promote professional self-improvement;¹³² (2) a pilot designed to collect information on magistrate judges to provide feedback during their terms and provide information to the relevant decision-makers when the magistrate judge seeks reappointment; and (3) a program that employs an independent commission to review a wide range of data on the performance of district, magistrate, and bankruptcy judges, and to distill that information into a written report describing each judge's strengths and weaknesses on the bench. These proposed pilots are described in more detail below, although the opportunity for variation extends far beyond those three. Upon completion, we envision that each of the pilot studies would themselves be evaluated to determine their value in providing useful feedback to judges, increasing transparency and process accountability in an appropriate manner, and promoting greater public understanding of the courts.

Pilot programs should be designed to test the application of different elements of state JPE programs to the federal arena. These elements include the types of information collected, how the information is collected, whether the information is provided directly to the judge or reviewed first by a supervisor or commission, whether specific recommendations concerning the judge's professional strengths and weaknesses are made, to whom evaluation results will be provided, and the specific goals of the program. We recognize that the ideal pilot studies would be controlled experiments; however, given the reality of the federal courts'

131. See INSTITUTE FOR THE ADVANCEMENT OF THE AMERICAN LEGAL SYSTEM & LEAGUE OF WOMEN VOTERS OF COLORADO, 2007 COLORADO VOTER OPINIONS ON THE JUDICIARY 4 (2007), available at <http://www.du.edu/legalinstitute/form-voter-input.html>.

132. In restricting these pilot proposals to the district court level, we do not mean to suggest that other federal judges, such as appellate judges and ALJs, should themselves have no formal evaluation. Indeed, at the state level appellate judges have been evaluated for decades, and thoughtful programs have been developed to tailor appellate evaluations to the specific tasks and responsibilities of those on the appellate bench. New programs for appellate judges are being considered as well. Among them, the State of New Hampshire recently established an internal committee responsible for developing individualized performance evaluations of its Supreme Court. See Maclean, *supra* note 19.

dockets and the concern about installing an untried JPE program at the federal level, much less rigorous approaches will have to suffice.¹³³

The three pilot studies described below build one on the other. The first proposed pilot is designed primarily to test the development of surveys, collection of meaningful case management data, and judicial response to receiving anonymous feedback. The second pilot includes each of these features, and additionally examines the value of interviews, independent review of judicial orders for clarity of communication, and distribution of evaluation results to those specifically charged with reappointing federal magistrate judges. The third proposed pilot would test the perceived advantages and disadvantages of an independent evaluation commission, and further would examine the efficacy of direct courtroom observation and judicial self-evaluations. While there is a logical progression to these proposals, they are not the only possibilities, and we welcome further discussion of the precise development of such pilots.

A. Pilot Proposal Number One: Confidential Evaluations

As the 1991 study in the Central District of Illinois suggested, a basic JPE pilot could be organized and conducted solely within the judicial branch at relatively low cost. Like the Central District of Illinois pilot, our first proposed program would preferably be piloted in a district in which all judges support the endeavor. Also like the Illinois pilot, the program would be based primarily on survey responses from those who have directly interacted with the judge in the courtroom. We would go beyond just attorneys who have appeared before the judge, however, and also issue surveys to litigants, court staff, and jurors where appropriate. For purposes of the pilot program, it may make sense to have these surveys developed by the Federal Judicial Center.

Survey data traditionally have formed the backbone of judicial performance evaluations, and some background on their use is warranted. Attorney surveys in particular already comprise the core of JPE programs at the state level. In addition, the Chicago Council of Lawyers has conducted survey-based evaluations of federal judges who sit in Chicago for over thirty-five years, including evaluations of magistrate judges, district judges, senior district judges—and judges on the Seventh Circuit Court of Appeals.¹³⁴ The most recent survey—seeking input on Chicago-area federal magistrate judges—was sent to approximately 3,400 members of the Federal Trial Bar for the Northern District of Illinois, as well as all members of the United States Attorney's Office and Federal Defender's

133. For a discussion of the benefits of controlled experimentation in the context of procedural rules, see Maurice Rosenberg, *The Impact of Procedure-Impact Studies in the Administration of Justice*, 51 LAW & CONTEMP. PROBS. 13, 14-16 (1988).

134. See CHICAGO COUNCIL OF LAWYERS, AN EVALUATION OF UNITED STATES MAGISTRATE JUDGES I (2008).

Office in Chicago.¹³⁵ Survey recipients were asked about the magistrate judge's integrity (for example, by indicating their level of agreement with the statement "His/her rulings in civil cases are free from any predisposition to decide for either plaintiff or defendant"), judicial temperament (e.g., "He/she is courteous toward lawyers and litigants"), legal ability (e.g., "He/she understands the issues in complex cases"), decisiveness (e.g., "He/she rules promptly on pretrial civil motions"),¹³⁶ and diligence (e.g., "His/her hearings and pretrial conferences reflect adequate research and preparation").¹³⁷

As the Chicago surveys suggest, attorney surveys must be carefully tailored both in their design and in their dissemination.¹³⁸ The survey should contain questions designed to elicit attorneys' perceptions of the judge's level of preparedness, clarity of expression, impartiality, and temperament on the bench, but should never allow them to comment on the substantive merits of a decision or order. With respect to dissemination, care must be taken to target only those attorneys who have actually appeared before the judge during the evaluation period, and to require those attorneys to respond to the survey based only on their personal experience with the judge.¹³⁹ In addition, an emerging practice at the state level is to survey attorneys shortly after the close of each case rather than to survey all attorneys at the same time.¹⁴⁰ Continuous dissemination of surveys has the virtue of targeting respondents while their experience with the judge is fresh in their minds.

Surveys should also invite attorneys to provide more extensive comments on the judge. In the pilot program in the Central District of Illinois, several of the judges who participated in the program reflected that the surveys and comments were beneficial to their professional development going forward. One district judge explained:

I have benefited from knowing the feelings, ratings, and views of the attorneys. We all develop habits or ways of doing, or not doing, things in connection with our offices that we often are

135. *Id.* at 2.

136. *Id.*

137. *Id.* at Exhibit 2.

138. See Steven Flanders, *Evaluating Judges: How Should the Bar Do It?*, 61 JUDICATURE 304, 304-05 (1978) (praising the efforts of the Chicago Council of Lawyers but cautioning that bar polls alone may not produce a fully accurate picture of the judiciary).

139. The State of Alaska actually permits attorneys to complete surveys based on professional reputation or social contacts with the judge, but they must clearly indicate that this is the basis for their answers, and the basis for such responses is noted in the judge's final evaluation. See UNIVERSITY OF ALASKA ANCHORAGE BEHAVIORAL HEALTH RESEARCH AND SERVICES, ALASKA JUDICIAL COUNCIL RETENTION SURVEY 1 (2004). No other jurisdiction permits evaluations based on anything other than direct experience.

140. See, e.g., COLO. RULES GOVERNING COMM'NS ON JUD. PERF. 10(a) (2007) (requiring that "surveys shall be conducted on a continuing basis").

oblivious to that need continuing or changing. The responses I got will aid me in doing my job.¹⁴¹

Similarly, the Colorado judges survey indicated that most judges appreciated the greater feedback and were able to translate that feedback into immediate positive change.¹⁴² As one judge put it:

The most useful part of the process is the survey results. Although I think we're never as good as the most glowing compliments and never as bad as the worst, it is sometimes possible to find a common thread that alerts you to deficiencies. Even the most hateful comments may contain a kernel of truth.¹⁴³

Attorney surveys are a necessary component of judicial evaluations, but other groups may also offer valuable information. Whereas attorney surveys in most jurisdictions ask for the attorney to rate a judge on a sliding scale for each question, juror surveys tend to ask a limited number of straightforward yes or no questions. This approach has the dual advantage of being easy to understand (yes or no questions make difficulties in the interpretation of questions less likely) and easy to complete (making it more likely that jurors will give it due attention after a long trial). As with attorney surveys, juror questionnaires tend to focus on the judge's behavior and control in the courtroom, rather than any substantive matter in the case. To this end, juror surveys might include questions such as: Did the judge treat people with courtesy? Did the judge act with patience and self-control? Did the judge act with humility and avoid arrogance? Did the judge pay attention to the proceedings throughout? Did the judge clearly explain the responsibility of the jury? Did the judge start court on time? And did the judge maintain control over the courtroom?¹⁴⁴

In practice, juror surveys tend to be overwhelmingly positive for judges.¹⁴⁵ This is clearly a good thing from the perspective of juror service and public perception of the courts. Jurors who leave a courtroom believing that the judge acted thoughtfully, fairly, compassionately (where appropriate), and with a firm hand are more likely to think about the courts as a steady and valued institution, and are also more likely to share their positive experience with others. As one commentator has put it, "The [jury] system has served many purposes, but its enduring purpose has been to secure a greater measure of trust in judicial institu-

141. DAVIS, *supra* note 50, at 9.

142. THE BENCH SPEAKS, *supra* note 79, at 13.

143. *Id.* at 14.

144. For related model juror questions, see SHARED EXPECTATIONS, *supra* note 8, at Appendix F.

145. See, e.g., STATE OF UTAH, UTAH VOTER INFORMATION PAMPHLET—GENERAL ELECTION NOVEMBER 2, 2004 at 46-69 (2004) (showing that jurors gave favorable responses of 95% or higher to virtually all survey questions for virtually all district judges evaluated in 2004).

tions.”¹⁴⁶ Because these virtues are so important, one should resist any temptation to discount high across-the-board ratings in juror surveys. Indeed, cause for concern should stem not from high ratings, but from abnormally low ones. The judge who does not connect with the jury, and who does not win jurors’ confidence or respect, fails in at least that aspect of his or her role as a public servant.

Litigants might also be surveyed, preferably shortly after the close of the case. The existing literature suggests that litigants’ satisfaction with the litigation process is far more relevant to their ultimate perception of the courts than the final outcome. Based on a review of court user studies in the 1980s, New York University Professor Tom Tyler concluded that “[t]he important role of procedural justice in mediating the political effects of experience means that fair procedures can act as a cushion of support when authorities are delivering unfavorable outcomes.”¹⁴⁷ The existence of litigant surveys also serves as a gentle reminder to judges that the perception of procedural fairness is as essential to the courtroom experience as the reality. As one judge put it, courts must be known for “fairness and respect, attention to human equality, a focus on careful listening, and a demand that people leave our courts understanding our orders.”¹⁴⁸

Nevertheless, because individual litigants tend to have the highest emotional investment in a case, surveys should be crafted with particular care to focus only on general aspects of the litigation process. Appropriate questions for litigant surveys may include: was the judge well-prepared for your case? Was the judge respectful to you? Were the judge’s rulings clear? Did the judge explain his or her ruling in a way that you could understand? And did the judge listen to your side of the case?¹⁴⁹

Self-represented litigants present an additional challenge, because they lack the mediating force of an attorney to help explain procedures and decipher rulings. And while family law matters—almost exclusively the province of the state court system—tend to see unrepresented litigants in particularly high numbers,¹⁵⁰ the growing costs of legal services means that federal courts are not immune from increasing numbers of pro se litigants. Indeed, over twenty thousand cases were filed by non-

146. Paul D. Carrington, *The Civil Jury and American Democracy*, 13 DUKE J. COMP. & INT’L L. 79, 93 (2003).

147. TOM R. TYLER, *WHY PEOPLE OBEY THE LAW* 107 (1990).

148. Kevin S. Burke, *A Court and a Judiciary that Is as Good as Its Promise*, 40 CT. REV. 4, 6 (2003).

149. See SHARED EXPECTATIONS, *supra* note 8, app. G.

150. In Colorado, nearly 56% of litigants in domestic relations cases were self-represented in 1999, and the percentage was growing. GOVERNOR’S TASK FORCE ON CIVIL JUSTICE REFORM, FINAL REPORT 35 (2000) (on file with authors), available at <http://www.state.co.us/cjrtf/report/report.htm> (last viewed Aug. 3, 2008).

prisoner pro se litigants in the federal courts in 2007.¹⁵¹ This does not mean, however, that pro se litigants require a different set of questions from those who do have legal representation. The survey instrument should include a question as to whether the respondent was self-represented, so that any trends concerning the judge's treatment of pro se litigants can be acknowledged.

Finally, surveys might be developed for court staff: law clerks, administrative staff, division clerks, court reporters, and others who interact with the judge on a regular basis in the courtroom. Such surveys should focus on the judge's interactions with support staff and clerks, level of preparedness in the courtroom, and responsiveness to administrative concerns, including case management issues.

For each of the respondent categories described above, care should be taken to ensure the anonymity of the respondents. Nothing on the survey should require or otherwise encourage the respondent to identify him or herself by name or specifics of the case. Where comments are provided, anonymity can be more thoroughly protected by asking someone unaffiliated with the judge being evaluated to review the comments and remove any identifying information before such comments are forwarded to the judge.

The number of responses to each survey is also important.¹⁵² Juror response rates can be kept high by requesting (or even requiring) that each juror complete the short survey at the end of trial as the final component of jury service. Attorney response rates, which are traditionally low, can be raised to adequate levels in two ways. First, the surveys might be sent out on a rolling basis, shortly after the termination of each case,¹⁵³ so that the specifics of the case and the judge's performance are fresh in the attorney's mind. Second, wherever possible, surveys should be sent out electronically. Recent developments in electronic survey software provide respondents with the same guarantees of anonymity, allow for surveys that can be completed quickly with a few mouse clicks, and make it easy for the survey provider to track the number of surveys sent out and returned.¹⁵⁴ Furthermore, the federal district courts' move to electronic filing in recent years means that virtually every attorney of

151. ADMINISTRATIVE OFFICE OF THE U.S. COURTS, 2007 ANNUAL REPORT OF THE DIRECTOR: JUDICIAL BUSINESS OF THE UNITED STATES COURTS, tbl. S-23 (2007).

152. See LAWRENCE F. LOCKE ET AL., READING AND UNDERSTANDING RESEARCH 48 (1998) (explaining the importance of an adequate sample size).

153. Here we define "termination" to mean the formal closing of the case, notwithstanding the possibility of appeal or reopening under other circumstances. Cases that close before an answer or other responsive pleading is filed would not be included.

154. See Gabriel M. Gelb & Betsy D. Gelb, *Internet Surveys for Trademark Litigation: Ready or Not, Here They Come*, 97 TRADEMARK REP. 1073, 1076-79 (2007) (describing the growth and development of online surveys); Dwight B. King, Jr., *User Surveys: Libraries Ask, "Hey, How Am I Doing?"*, 97 LAW LIBR. J. 103, 109-112 (2005) (same).

record has a valid e-mail account on file with the court, making accurate outreach to attorneys a relatively simple matter.¹⁵⁵

Response rate and anonymity are more significant challenges for litigant surveys, although those challenges are not insurmountable. Most litigants (other than those proceeding pro se) do not automatically provide a physical or e-mail address to the court. Litigants may also feel less of an obligation to complete the survey than attorneys or jurors. Response rates can be increased, however, by making the opportunity to provide feedback easily available. The use of comment cards in restaurants provides an analogy. As one recent study concluded:

If a customer has to seek out a comment card from a host or hostess, the cashier, or the front desk, suggestions for improvement of operations and general customer feedback on service are likely not to be received. Instead, relatively infrequent comments relating to extreme situations will likely be the only feedback provided. Although useful to know about these situations, it is of much greater importance to continually have the typical customer's assessment of normal operating conditions.¹⁵⁶

One method of creating a larger and more representative response rate is to make it easy for each litigant to provide information by providing the litigant with information at the end of the case that identifies a court website and gives a specific (and time-limited) password to log into the system. The respondent could use any computer to complete the survey. Another possibility is to develop kiosks in the courthouse that would allow litigants to complete evaluation surveys electronically after leaving the courtroom. Of course, survey questions would have to be carefully formulated to protect against responses based on the emotion of an immediate courtroom appearance. But the technology for such kiosks already exists, and is being put to use in several state courthouses.¹⁵⁷

Beyond survey data, we propose that the first pilot include collection of individual judges' case management data. Such data could be compiled by the Administrative Office of the U.S. Courts and sent to the judge to introduce an additional perspective on the judge's overall per-

155. E-mail has become so essential to electronic case filing that one recent decision suggests that an attorney's failure to check the status of a case via e-mail or the PACER system may constitute professional malpractice. See Jessica Belskis, *Electronic Case Filing: Is Failure to Check Related to an Electronically Filed Case Malpractice?*, 2 SHIDLER J.L. COM. & TECH. 13, 13 (2005) (discussing *Blackburn v. U.S. Dep't of Agric. & Forest Serv.*, No. C04-1404RSM (W.D. Wash. 2005)).

156. Joel D. Wisner & William J. Corney, *An Empirical Study of Customer Comment Card Quality and Design Characteristics*, 101 BRITISH FOOD J. 621, 629 (1999).

157. See Randall T. Shepard, *The New Role of State Supreme Courts as Engines of Court Reform*, 81 N.Y.U. L. REV. 1535, 1546 (2006) (discussing the emergence of "pro se kiosks" in Maricopa County, Arizona); Henri E. Cauvin, *New Internet Kiosks Make Courts More User-Friendly*, WASH. POST, Jan. 18, 2007, at DZ03 (noting installation of kiosks in the D.C. Superior Court to provide information and allow payment of fines).

formance. In addition to data concerning closed cases, average time to disposition, average caseload, and the like, at the pilot court's election, the data might also indicate to each judge where his or her statistics rank within the district, the circuit, and the nation. We float this idea not because we believe that judges should be ranked crudely against each other, but rather because relative performance, particularly within a district where judges have comparable dockets, is another important piece of information.

Judges in the 2008 Colorado survey expressed overwhelming support for the inclusion of case management data in performance evaluations. Overall, 73% of trial judges surveyed agreed or strongly agreed with the statement that case management data should be part of the evaluation process.¹⁵⁸ The judges in the survey expressed caution that such data be considered carefully and thoughtfully, and that judges be given the opportunity to explain to the committee any unusual issues with their dockets.¹⁵⁹ A federal JPE pilot might strike the same balance in the use of case management data; such data could be reviewed carefully for each judge and afforded appropriate, but not undue, weight in light of all the other factors comprising the judge's performance.

As envisioned in this pilot program, formal "evaluations" would not exist in a concrete sense. Rather, the survey and case management data, and relevant survey comments, would be sent to each judge for review and consideration toward his or her professional self-improvement. A slightly more robust version of this pilot would use the survey and case management results for individual mentoring or collective judicial training sessions.

At the completion of the pilot program, a separate study should be conducted to determine whether (and to what extent) the program met its stated goals, and which aspects of the program are worth maintaining, developing further, or discarding. This study could be conducted by the FJC or an organization unaffiliated with the federal judiciary that has similar capacity to conduct high-quality analysis.¹⁶⁰ Any such study should seek to measure all data reasonably related to the pilot program's goals. Here, the goals would include developing useful and appropriate survey instruments, soliciting an adequate number of survey responses, and providing meaningful information to judges in the form of survey and case management data. Much of the data that can be collected is

158. THE BENCH SPEAKS, *supra* note 79, at 14.

159. *See id.* at 15-19.

160. Assuming adequate resources are available, the benefits to using the FJC are rather obvious. The FJC's knowledge of the federal courts and the circumstances under which they operate makes it a natural first choice. At the same time, however, using a competent organization outside the judiciary to review the effectiveness of the pilot projects would remove any charge that the federal courts were simply reviewing themselves and might add to public confidence in the conclusions ultimately reached.

objective: how many judges were evaluated? Did all judges participate voluntarily? How frequently did evaluations take place? What criteria were set for evaluation? Other effectiveness data are subjective but highly important: did the judges find the feedback they received to be useful? Did judges express any concerns about the impact of the program on their decisional independence? Did the process generate any evidence of increased confidence among the bar or the public or both?

We recognize that the most important goal, improved judicial performance, would be difficult to measure meaningfully in the context of a pilot program. Careful selection of measurement variables, however, may help approximate at least the perceptions of improved performance. This approximation can be further strengthened by the inclusion of additional structural and data elements in the pilot program. We turn next to an example of this type of more robust pilot—one that adds interviews, review of written orders, and supervisory feedback.

B. Pilot Proposal Number Two: Magistrate Evaluations

A second proposed pilot program would be limited to the evaluation of magistrate judges, but would be expanded with respect to the amount of information collected and the purposes for which such information was used. Specifically, the program would be designed to provide constructive feedback to magistrate judges well before the time they would need to announce intent to seek reappointment, and also to provide the historical evaluations of each magistrate judge to those vested with the responsibility of making reappointment decisions.

The value of comprehensive evaluations to the reappointment process is a primary goal of this proposed pilot. Currently, the information available for reappointment recommendations is relatively limited. Assuming the district court decides to consider an incumbent for reappointment, an independent panel is created to collect information and recommend whether the magistrate judge should be awarded an additional term.¹⁶¹ The panel typically seeks input on the magistrate's performance through public comments, although an interview with the incumbent is also encouraged.¹⁶² But as Professor Resnik has documented in the context of bankruptcy judge reappointments, giving heavy weight to public comments poses considerable risk. Comments come from self-

161. At least one year before the expiration of the magistrate judge's term, the district court must inform the magistrate judge whether it has determined not to reappoint the magistrate judge, or whether it has determined to consider reappointing the magistrate judge. ADMINISTRATIVE OFFICE OF THE UNITED STATES COURTS, THE SELECTION, APPOINTMENT, AND REAPPOINTMENT OF UNITED STATES MAGISTRATE JUDGES 29 (2002). If the district court chooses the latter route, it must issue a public notice soliciting volunteers to serve on a "merit selection panel" for reappointment. *Id.* at 30. The panel must include at least two lawyers and at least two non-lawyers. *Id.* at 12.

162. *Id.* at 30.

selected respondents who may have particular agendas.¹⁶³ Furthermore, lawyers whose critical comments would otherwise be valuable may be reluctant to comment for fear that, should the judge be reappointed notwithstanding their feedback, they may have to appear before the judge with their identities—and criticisms—known.¹⁶⁴ The proposed pilot would test whether these concerns could be alleviated by broadly expanding the pool of information available to the reappointment panel, as well as employing anonymous surveys administered under the same rigorous methodology as proposed in the first pilot program.

As with the first pilot program described above, the magistrate pilot would begin with the collection of survey data and relevant case management data. In addition, the developers of the pilot program might consider including two additional sources of information: an interview with the magistrate judge, and a sample of the magistrate judge's recent written orders.¹⁶⁵

Reappointment panels already have the option to interview an incumbent magistrate judge.¹⁶⁶ In the pilot program, we would suggest using an interview to flesh out any concerns about the collected information and allow the magistrate judge to provide any additional information that might not be evident from the data, such as an irregular docket or an unusually demanding or notorious case. The interview might also preview the panel's perceptions of the magistrate judge's strengths and weaknesses, and foster discussion (at least at a basic level) about future efforts for continuous professional improvement. Accordingly, the optimal time for an interview is some time after the panel has collected survey and case management data. All collected information should be forwarded to the magistrate judge in advance of the interview so he or she has an opportunity to review it ahead of time.

This pilot program might also consider collecting samples of the magistrate judge's orders for review. The purpose of this review is two-fold. First, it allows those charged with reappointment to determine whether the magistrate judge's orders and opinions are sufficiently clear and understandable. It goes without saying that when any judge expresses him or herself in writing, the attorneys and litigants who review the opinion should understand without any hesitation the precise scope of the judge's ruling, and how the judge reached that result. Put another way, the parties should be clear about what happens next in their case.

163. Judith Resnik, "Uncle Sam Modernizes His Justice": *Inventing the Federal District Courts of the Twentieth Century for the District of Columbia and the Nation*, 90 GEO. L.J. 607, 676 (2002).

164. *Id.*

165. While discussed here solely in the context of magistrate judges, self-evaluations and review of written orders could obviously be piloted, and, if successful, employed in any JPE program for federal district judges, bankruptcy, judge, and the like.

166. ADMINISTRATIVE OFFICE OF THE UNITED STATES COURTS, *supra* note 161, at 30.

But this concern goes beyond the parties; even those with no interest in a particular case outcome should be able to read an opinion and understand the judge's reasoning.

The second purpose behind reviewing written orders is to ensure that the magistrate judge is relying on sound legal reasoning. By this we do not mean a "correct" decision in the substantive sense—that determination is the province of the district or appellate courts or both. Rather, the review for JPE purposes is somewhat more high-level: has the magistrate judge provided adequate citation to legal authority? Did the magistrate judge set out the relevant facts and evidence on which he or she relies for the decision? And is there a logical flow to the reasoning in the opinion?

While we have described this pilot with specific reference to magistrate judges, it could apply with limited modifications to the evaluation of bankruptcy judges as well. Reappointment decisions concerning bankruptcy judges, for example, do not require the use of advisory panels, but the practice of providing evaluation information to the ultimate reappointment authority would remain the same.

As with the first pilot, a pilot program using this model should be evaluated for its effectiveness. Because this proposed model builds on the first pilot, each of the effectiveness measurements used for the first pilot are equally appropriate here. In addition, one might examine the effectiveness of this pilot study by attempting to measure the value of a face-to-face interview to both the magistrate judge and the reappointment panel, the value to the reappointment panel of reviewing the magistrate judge's sample of written orders, and the overall value of the collected information on the magistrate judge to the reappointment panel.

C. Pilot Proposal Number Three: Commission-Based Evaluations

A final proposed pilot program—again, one of dozens of possible variations—combines the emphasis on broader data collection in the proposed magistrate judge pilot with an additional element frequently used in state JPE programs: an independent commission to conduct the evaluation. The inclusion of an independent commission may serve three purposes. First, commission members may provide a perspective on the judge's professional strengths and weaknesses that might not be directly apparent to the judge, even if both the judge and the commission are reviewing the same performance-related information. Second, the commission can process the collected information and summarize it in a form that is more easily digestible than survey and case management data, and a potentially unwieldy number of comments. Finally, and perhaps most importantly, an independent commission may provide an imprimatur of balance and public oversight that simply cannot be achieved by even the most conscientious internal evaluation.

By “independent commission” we mean a volunteer commission at the district court level composed of a roughly equal number of attorneys and non-attorneys. This balance has proven to be successful at the state level,¹⁶⁷ and a similar format is used for magistrate judge reappointment panels.¹⁶⁸ Attorneys have the most consistent direct experience with the court and should have reasonable expectations about the nature of the judicial process. Attorneys may also provide legal expertise during the evaluation process and are in a position to explain the intricacies of a court’s docket to laypersons. Non-attorneys provide a community perspective in the evaluation process, and help assure that the final evaluations are presented in a straightforward and non-technical manner. Both groups also add to the ultimate legitimacy of the evaluations. If only attorneys were involved, the process might be criticized as an “inside job” in which members of the legal community simply protect their own.¹⁶⁹ Citizen involvement strengthens both the perception of the program and the final product. As one early commentator on JPE put it:

The key to any successful program of judicial evaluation is active lay participation—people working in concert or as a part of a co-ordinated effort with the legal profession in a broadly based citizens’ effort to assist the voters in making those important decisions on critical judicial positions.¹⁷⁰

If no attorneys were involved, however, evaluations might equally be criticized as the product of those who have little or no familiarity with the legal system or the role of judges. Finally, the inclusion of non-attorneys carries the additional benefit of fostering greater community understanding of the role of the judge. At least one study has shown that non-attorneys who are involved in the evaluation process walked away from it with a better sense of what judges do.¹⁷¹

For purposes of testing a JPE program in a pilot setting, it is sufficient to assemble a dedicated and balanced group of ten to twelve volunteers who are willing to review the relevant information provided by the

167. As of 2008, all seven states with comprehensive JPE programs (Alaska, Arizona, Colorado, Kansas, New Mexico, Tennessee and Utah) utilize a commission with roughly equal representation of attorneys and laypersons.

168. See ADMINISTRATIVE OFFICE OF THE UNITED STATES COURTS, *supra* note 161, at 12.

169. See A. John Pelander, *Judicial Performance Review in Arizona: Goals, Practical Effects and Concerns*, 30 ARIZ. ST. L.J. 643, 648-49 (1998); John M. Roll, *Merit Selection: The Arizona Experience*, 22 ARIZ. ST. L.J. 837, 878-79 (1990). The same criticisms have emerged where evaluations are conducted by Judicial Councils or otherwise controlled by the judiciary itself. The State of Utah, for example, instituted a JPE program in 1984 under the auspices of its Judicial Council, which consists of twelve judges and one attorney. That approach came under fire in the mid-2000’s and was a major consideration undertaken by the state’s Judicial Retention Task Force in 2007. In March 2008, Utah passed new JPE legislation that, among other things, entrusted evaluations to a commission composed of attorneys and non-attorneys. See UTAH CODE ANN. § 78A-12-201 (2008).

170. Henry T. Reath, *Judicial Evaluation—The Counterpart to Merit Selection*, 60 A.B.A. J. 1246, 1247 (1974).

171. See, e.g., Anne Rankin Mahoney, *Citizen Evaluation of Judicial Performance: The Colorado Experience*, 72 JUDICATURE 210, 216 (1989).

Clerk of the Court (i.e., the survey data, self-evaluation, and sample of written orders) and the Administrative Office of the U.S. Courts (case management data), collectively discuss the judge's strengths and weaknesses that come out in the collected information, and draft a short report describing their conclusions as to the judge's performance. The independent commission may also wish to collect information itself from two additional sources: courtroom observation and a judicial self-evaluation.

Courtroom observation may be performed by members of the independent commission or by additional courtroom observers specially trained for that purpose.¹⁷² There are advantages to either approach. Where commission members themselves conduct the observations, they can witness directly the judge's courtroom behavior and the behavior the judge expects (and tolerates) from attorneys, witnesses, and jurors. Commission members who do their own observations are also more likely to internalize the challenges a judge may face in the courtroom—crowded dockets, emotionally distraught litigants, unprepared counsel, and so on—that may not be apparent from paper reports on the judge. Independent observers, by contrast, are less likely to be recognized by the judge (assuring that the judge's behavior is not altered by the knowledge that a commission member is in the courtroom), and may have fewer preconceived notions about the judge before entering the courtroom.

In the Colorado judges survey, trial judges overwhelmingly supported regular courtroom observation. Forty-eight percent of the judges stated that courtroom observation was “very useful” in the JPE process, and another 40% stated that such observation was “somewhat useful.”¹⁷³ Only 12% were neutral on the issue, and no judge indicated that observation was not useful.¹⁷⁴ Indeed, the judges practically pleaded for more observation as part of the process,¹⁷⁵ noting among other things that ob-

172. The State of Alaska, for example, has used a special corps of courtroom observers who are trained in advance and are required to couch their observations in specific categories of predetermined, process-oriented criteria. As many as fifteen observers are assigned to each judge. Each observer is given approximately forty hours of advance training, and the observers are directed to sit in courtroom proceedings at unscheduled intervals. They observe both criminal and civil cases and proceedings ranging from arraignments and motion hearings to full jury trials. The observers' notes are collected into a report for each judge, which specifies the number of observations, types of events and cases observed, the total number of hours the judge was observed, and the average rating the judge received in each category. The final reports are then forwarded to the Alaska Judicial Council for consideration as part of the judge's overall evaluation. See ALASKA JUDICIAL OBSERVERS 2006 BIENNIAL REPORT 1-8 (on file with authors), available at <http://www.ajc.state.ak.us/Retention2006/JudicialObservers2006.pdf>.

173. THE BENCH SPEAKS, *supra* note 79, at 20.

174. *Id.*

175. Colorado's JPE program requires that each commission member directly observe at least three judges up for evaluation in unannounced courtroom visits. See COLO. REV. STAT. § 13-5.5-103(1)(k) & -105(1)(c) (2008); Jane B. Howell, Executive Director, Colorado Office of Judicial Performance Evaluation, Presentation on Colorado Commissions on Judicial Performance (Aug. 2008), http://www.courts.state.co.us/Media/Law_School.cfm (select link for “Judicial Performance and Retention Presentation”) (last visited Oct. 21, 2008).

servation would be enhanced further if the observer were to meet with the judge immediately afterward to ask any questions about what occurred. One judge summarized the feelings of a majority of respondents as follows:

Observation is particularly useful if the judge does not know the individual or does not know that the observer is present in the courtroom. In our district, a courtroom can be observed by security officers over closed circuit TV. Another way is to appear during a busy docket day when the courtroom is full. I think the observer should arrange to sit down with the judge after observing him or her and ask questions about things observed to be sure there is no misunderstanding—particularly if the observer is a lay person.¹⁷⁶

The developers of the pilot program may also consider introducing a judicial self-evaluation. It should be designed with two purposes in mind: to allow each judge to consider his or her strengths and weaknesses on the bench in private reflection, and to help the independent commission determine whether the judge's perception of those strengths and weaknesses comport with the strengths and weaknesses identified by others. The first purpose recognizes that even the most conscientious judge committed to continuous professional improvement is more likely to succeed when he or she takes time to assess his or her skill set on a regular basis. The second purpose ensures that there is no disparity between the judge's perception of his or her skills (positive or negative) and the perception of those who interact with the judge in the courtroom.

A common complaint among judges is that self-evaluation lacks real meaning, because all judges feel pressure to rate themselves as highly as possible.¹⁷⁷ Indeed, an average or below average self-evaluation in any category may invite questions from the commission, even if all other data indicate that the judge is above average in that category. There is, of course, little that can be done to assure that all judges approach the self-evaluation openly and honestly. But those who do should derive a benefit much greater than those who are inclined to inflate their self-assessments.

The pilot program can help promote sincere self-evaluations by considering making available mentorship programs that foster self-improvement. Here, the State of Arizona provides a model. Each judge who is evaluated is assigned to a "conference team" composed of another judge, a member of the state bar, and a member of the public.¹⁷⁸ The judge meets with the conference team to discuss his or her strengths,

176. THE BENCH SPEAKS, *supra* note 79, at 21.

177. *Id.* at 21-22.

178. See Pelander, *supra* note 169, at 690.

weaknesses, and areas for improvement based on the self-evaluation, survey results, and public comments. Together, the judge and his or her conference team prepare and sign a written self-improvement plan.¹⁷⁹ By rule, no member of the conference team may participate in formal evaluation of the judge.¹⁸⁰

Similar to Arizona, a federal JPE pilot program might choose to use judicial self-evaluations as a teaching and self-improvement tool rather than as a formal component of the magistrate judge's evaluation. The conference team model, which uses the information gleaned from the evaluation process but is more hands-on and more private than a formal evaluation, should encourage magistrate judges to be thoughtful and honest in their self-assessments.

Regardless of whether the judge being evaluated is a district judge, magistrate judge, or bankruptcy judge, the independent commission should endeavor to collect as much of the specified information as possible during the data collection phase. If all information is not available for a certain judge, however, the commission may still proceed with the evaluation and note in its report what information was missing.

Once again, an effectiveness study should follow a pilot program of this type. In addition to the measurements described for the first and second proposed pilots, an effectiveness study here should consider measuring the time and money expended by the independent commission, the commission members' perception of the process, the value each judge placed on completing the self-evaluation, the number of times each judge was observed in the courtroom, the value to both the judge and the commission of the courtroom observation, and the overall value to the judge of receiving an independent analysis of his or her performance data.

IV. BEYOND PILOT PROGRAMS: CHALLENGES AND OPPORTUNITIES

It is our hope that suitable testing of JPE through a variety of pilot programs in one or more districts will more rigorously define the contours of successful judicial performance evaluation at the federal level. A successful pilot program in one district, however, may not translate cleanly to another district, or to a national program encompassing all ninety-four districts. A pilot program may also be able to sidestep certain structural and administrative issues that become more pressing—and more complex—when extrapolated to a national program. The development of a more permanent and expansive federal JPE program, then, ultimately will require consideration of several additional factors. Here we discuss five such factors: the administration of a national JPE pro-

179. *Id.* at 692; *see also* ARIZ. JUD. PERF. REV. R. 4(f).

180. *See* Pelander, *supra* note 169, at 693.

gram; the development of national rules governing such a program; the structure and use of independent commissions; the frequency of evaluations; and the degree to which evaluation results should be publicly disseminated.

A. Authorization and Administration

A nationwide program of JPE in the federal courts would plainly benefit from some degree of centralized administration. A single authority can exercise greater control over both inputs (e.g., uniform survey questions, case management data, and operating procedures) and outputs (e.g., format of evaluation reports, methods of disseminating evaluation results if applicable). A centralized authority also may find it easier to run comparative analyses across districts, to confirm (for example) that the level of survey responses is reasonably consistent.

Judges would presumably be more comfortable if any national administration of a federal JPE program rested within the judiciary itself. Excellent resources within the judicial branch for developing and operating a national program are already in place. In one formulation, the Judicial Conference of the United States might be tasked with overseeing the program and developing national rules and procedures; the FJC with developing appropriate surveys and conducting research into the effectiveness of JPE across districts; the Administrative Office of the U.S. Courts with providing case management data; and the Circuit Clerks with collecting all the relevant data for each evaluated judge in the circuit and providing that information to the judge, appointing authority, and independent commission as appropriate.

A coordinated effort within the federal courts to develop and implement a robust JPE program would require both commitment and leadership from the judiciary. But it would also assure greater judicial control over the process and would demonstrate to the American public that federal judges are sincere about improving their performance on the bench and committed to their role as public servants. By contrast, the potential consequence of a tepid judicial response is a strong—and perhaps unwelcome—legislative response. Titles 18 and 28 of the U.S. Code are littered with statutory responses to perceived failures of the federal judiciary to address issues invoking accountability. The Speedy Trial Act¹⁸¹ reflects legislative impatience with the judicial use of Federal Rule of Criminal Procedure 50(b).¹⁸² The Judicial Conduct and Disability Act arose in part as a response to the failure of Judicial Councils to use their (admittedly vague) authority to deal with issues of judicial mis-

181. See, e.g., 18 U.S.C. § 3161 (2008).

182. See Linda S. Mullenix, *The Counter-Reformation in Procedural Justice*, 77 MINN. L. REV. 375, 430 (1992).

conduct under the pre-1980 version of 28 U.S.C. § 332.¹⁸³ The Civil Justice Reform Act was in part a Congressional response to the perception that district courts were not sufficiently achieving cost and delay reduction through case management.¹⁸⁴ And the Alternative Dispute Resolution Act of 1998 suggests legislative frustration with the pace of alternative dispute resolution implementation in the federal courts.¹⁸⁵

When the judicial branch demonstrates strong leadership in areas of transparency and accountability, however, Congress is less likely to pass reactionary legislation. The Judicial Conference's mandated use of conflict disclosure software, for example, probably prevented a legislative push for a statutory mandate.¹⁸⁶ A similar mandate concerning seminar disclosure requirements may be sufficient to defeat a proposed legislative amendment to the pending federal judicial salary legislation to ban attendance at non-judicial seminars altogether.¹⁸⁷ And broad legislative demand for cameras in the courtroom in at least the lower federal courts may be dampened by the experimental posting of audio files of trial court proceedings on the federal court's PACER website.¹⁸⁸

Judicial leadership and judicial administration of a federal JPE program, then, may well make sense. A statutory structure, however, does not entirely lack merit. We note that in most states in which JPE is successful, it rests upon a legislative foundation. Moreover, thoughtful legislation may carry certain advantages. First, a JPE program may engender greater public confidence when it is not administered by the judiciary itself. No matter how well-meaning or thoughtfully constructed a judicially operated JPE program might be, from the citizen perspective the evaluations are to some degree less credible when judges are the only parties involved in their own evaluations.¹⁸⁹ Second, a statute provides backbone to a program; it may be somewhat more difficult to change than a set of rules or guidelines, and accordingly may provide a more

183. See Stephen B. Burbank, *Procedural Rulemaking under the Judicial Councils Reform and Judicial Conduct and Disability Act of 1980*, 131 U. PA. L. REV. 283, 291-94 (1982) (describing the legislative history of the Act and the concerns of some that the Judicial Councils were unwilling to exercise existing power).

184. See, e.g., Mary Brigid McManamon, *Is the Recent Frenzy of Civil Justice Reform a Cure-All or a Placebo? An Examination of the Plans of Two Pilot Districts*, 11 REV. LITIG. 329, 332 (1992).

185. 28 U.S.C. § 651 (2008).

186. The Third Branch, Administrative Office of the U.S. Courts, Washington, D.C. *Transparency Takes Shape* (Dec. 2006), available at <http://www.uscourts.gov/ttb/12-06/transparency/index.html>; see also Linda Greenhouse, *Federal Judges Take Steps to Improve Accountability*, N.Y. TIMES, Sep. 20, 2006, at A20.

187. See Greenhouse, *supra* note 186; see also Charles Lane, *Judges Alter Rules for Sponsored Trips*, WASH. POST, Sep. 20, 2006, at A23.

188. See News Release, U.S. Courts, Pilot Project Begins: Two Courts Offer Digital Audio Recordings Online (Aug. 6, 2007) (on file with authors).

189. See Jean E. Dubofsky, *Judicial Performance Review: A Balance Between Judicial Independence and Public Accountability*, 34 FORDHAM URB. L.J. 315, 334 (2007) (“[J]udges’ own evaluations often are too self-serving; no one can possibly be as good as some judges think they are.”).

consistent model for judicial performance than might otherwise be available. Individual rules and guidelines can of course be authorized by the legislative framework, and such guidelines can be adjusted with greater frequency while leaving the framework in place.¹⁹⁰ Third, a statute represents, at its best, a triumph of communication between two branches of government, and between the citizens and their courts. It should be a product of discussion between elected representatives, the courts, and the public, based on careful thought and discussion about what it means to evaluate judges as public servants.¹⁹¹ Finally, legislation would promote a certain degree of transparency about the purpose of the JPE program; debates concerning the legislation become part of the public record, and the statute itself is widely available. If the goals of the program and the criteria for evaluation are clearly delineated, the statute itself can be a self-contained civics lesson on what the public can expect from its federal judiciary.

We are cognizant of the concerns—voiced most recently by Justice Wells—that legislative authorization opens the door to legislative mischief.¹⁹² And in discussing the perceived advantages and disadvantages to legislative versus judicial authorization of a federal JPE program, we do not take a position that one approach is preferable. A robust program created and designed under the auspices of the judiciary, with active leadership and buy-in from the entire federal bench, would be expected to gain considerable public support and plaudits for the courts' commitment to transparency and accountability.

B. Rules and Procedures

Depending on the frequency and complexity of pilot JPE programs, rules and guidelines governing those programs may reasonably be set at the district or circuit level. Indeed, this sort of experimentation may be encouraged as a means of developing JPE rules and practices that best meet with the performance improvement needs of the judges and the

190. We discuss the ability to change or amend rules and guidelines in greater detail in Part IV(B) *infra*.

191. We fully recognize that the process of crafting legislation is inherently political and that even the most careful strategies and thoughtful suggestions are not immune from criticism and compromise. But that sober conclusion warrants greater involvement in the process, not less. Two recent legislative amendments to state JPE programs illustrate the point. In Utah, the Task Force on Judicial Retention considered changes to the state's program through the fall of 2007 before a bill was introduced in the state senate in early 2008. The Task Force was composed of members of both houses of the state legislature and three state court representatives, including the Chief Justice. See Minutes of the Judicial Retention Election Task Force (Aug. 14, 2007) (on file with authors). While it is not fair to say that every member of the Task Force walked away from the experience with everything he or she wanted, the discussion only benefitted the final legislative outcome. Similarly, the respective chairs of Colorado's Senate and House Judiciary Committees held open meetings throughout the summer and fall of 2007 to receive input on the reenactment of that state's JPE statute. The interests of the courts, attorneys' groups, legal organizations, think tanks and private individuals were considered during these initial meetings and throughout the legislative process. The bill that ultimately became law, while perfect to no one, was nevertheless largely acceptable to all.

192. See generally Wells, *supra* note 112 and accompanying text.

information-seeking needs of appointment authorities, chief judges, or the public. In a national program, however, a uniform set of national rules and guidelines should be considered, perhaps to be developed by the Judicial Conference. Such rules might set out the purposes of the JPE program, criteria for evaluation, process of collecting data, person(s) eligible to view the collected data on each judge, and procedures for preparing a report if an independent commission is used.

A provision for future amendment might also be included in the rules. Amendments should not go into effect immediately, for the obvious reasons that independent commissions would require time to prepare for changes to the data collection or the evaluation procedure, and judges subject to evaluation would require adequate notice of a procedural change. Judicial performance evaluations only have professional development impact if each judge is presented with clear and comprehensible criteria for the evaluation ahead of time. Put another way, judges cannot be expected to hit a constantly moving target.

C. Use and Composition of Independent Commissions

If pilot studies suggest that a national program should use independent commissions, the rules should set out the composition and reach of those commissions. As discussed above, for pilot purposes it may be sufficient to seek out volunteers in rough balance of attorneys and non-attorneys. In a more fixed program, however, consideration should be given also to balancing a commission by appointing authority, and perhaps to achieving partisan balance as well. The judiciary may wish to appoint all commissioners itself. Or the rules could institute a model in use in some states, in which appointments are shared among all three branches of government. Some appointment authority may also be invested in groups outside the federal government that have an investment in the success of the courts, such as the Faculty of Federal Advocates or a local bar association. No matter how appointments are broken down, additional balance can be achieved by staggering the terms of each commission member so that no appointing authority has the power to pack the commission with new appointees at any given time.¹⁹³

D. Frequency of Evaluations

A judge is likely only to be evaluated once or twice during the course of a pilot program. In a more permanent program, however, regular intervals should be set for the evaluation of each judge. Evaluations should be conducted with sufficient frequency that improvements and continuing strengths are well-documented, and weaknesses or difficulties are identified and diagnosed quickly. Depending on the stated goals of a

193. For a more extensive discussion of best practices for the composition of independent commissions at the state level, see SHARED EXPECTATIONS, *supra* note 8, at 81-82.

final JPE program, evaluations for magistrate and bankruptcy judges might be timed to provide reasonably current data to assist with reappointment decisions. For example, districts that employ part-time magistrate judges with four-year terms might evaluate those magistrate judges three years after their initial appointments and three years after each reappointment.¹⁹⁴

Full-time magistrate judges and bankruptcy judges might be evaluated more often. More frequent evaluation would serve two purposes. First, it would identify the judge's strengths and weaknesses on an early and continuing basis, providing more time and opportunity for professional development than might be the case if the judge was only evaluated once each term. Second, multiple evaluations are more likely to show improvement or lack of improvement in certain areas over time, making aberrations in data or dockets less likely than might be the case if only one evaluation is conducted.

The timing of evaluations for district judges with life tenure is less obvious, because there are no reappointment decisions or other temporal cues. However, there are strong arguments for maintaining a pattern of regular evaluations similar to those proposed for magistrate judges. One format might be to evaluate district judges two years after initially taking the bench, and every three years thereafter. District judges with senior status could also be evaluated every three years. This scheme has been used at the state level where judges are similarly appointed with no fixed terms. In New Hampshire, for example, Superior Court and District Court judges face evaluation at least once every three years.¹⁹⁵ Similarly, in Massachusetts, trial judges are evaluated at least once every three years.¹⁹⁶

Evaluation every three years makes sense from a professional development perspective. As is the case with magistrate judges, regular evaluation of district judges helps to identify and address weaknesses more quickly. Regular evaluation also can be used to demonstrate a pattern of individual improvement. It can pinpoint areas for more widespread judicial education if a number of evaluations suggest collective strengths and weaknesses across the court. Finally, when evaluations are conducted on a frequent and consistent basis, the natural level of stress associated with such evaluations (for the judge being evaluated, but also

194. This timing scheme would allow sufficient time for a thorough evaluation (and, if requested, a recommendation on reappointment) based on three years of the magistrate judge's performance. It would also afford the magistrate judge a full year before reappointment to address any weaknesses identified in the evaluation.

195. See News Advisory, *New Hampshire Judiciary, 2002 Judicial Performance Evaluations Released* (Jul. 11, 2003) (on file with authors) available at <http://www.courts.state.nh.us/press/pr030611.htm>.

196. MASS. GEN. LAWS. ch. 211, § 26A (2008) (setting evaluation intervals of twelve to eighteen months for judges with four years of experience or less, and evaluation intervals of eighteen to thirty-six months for judges with more than four years of experience).

for members of the bar and court staff who must interact with the judge on a regular basis) may be diluted.

E. Dissemination of Evaluation Results

Public dissemination of evaluation results historically has been a difficult aspect of JPE for judges to stomach. And there are certainly circumstances in which the release of a judge's evaluation may be unwise.¹⁹⁷ But the importance of circulating evaluation results to the general public should not be underestimated, and should be considered carefully for inclusion in any final program of judicial performance evaluation. Indeed, there are at least five advantages to broad sharing of results. First, regular dissemination of judicial evaluations reinforces appropriate criteria by which the public should measure its judges. Rather than considering a judge to be "good" or "bad" based on his or her handling of a particularly difficult or controversial case (the types of cases most likely to be covered by the media)—or worse, the outcome of a particularly politicized case—the lay public may begin to base its assessment of judicial quality on the judge's clarity of expression, ability to manage his or her docket efficiently, and demonstrated command of procedural rules and substantive law.

Second, to the extent the public is aware that evaluations are being conducted, knowledge of the program may increase transparency and foster public confidence in the court system. Transparency would in fact be increased along two dimensions: the courts demonstrate that they are willing to be subject to evaluation as a means of fostering continuous improvement in their role as public servants, and the evaluation commission demonstrates that its evaluation process was open, thorough, and in line with its role as a representative of the people.¹⁹⁸

197. If, for example, the judge retires or resigns after the evaluation is conducted but before the date scheduled for its release, there is little value in making the evaluation public. This approach has precedent at the state level. Colorado has a longstanding requirement that judges be shown their final evaluation and "narrative profile" (a short form of the evaluation for inclusion in voter guides) at least 45 days before they must declare their intent to seek retention. COLO. REV. STAT. §§ 13-5.5-106(1)(a)(V) & (2)(A)(3) (2008). Judges who resign from the bench or do not declare their intent to seek retention do not have their evaluations circulated or placed in the voter guide. In 2008, the state's JPE legislation was amended to require the state's Office of Judicial Performance Evaluation to issue a statewide statistical report thirty days before the election, setting forth the total number of justices and judges who were eligible to stand for retention, the total number of evaluations performed by the state and district commissions, the total number of justices or judges who were evaluated but did not stand for retention, and the total number of justices and judges recommended for retention. *See id.* §§ 13-5.5-103(1)(q)(I)-(IV).

198. One commentator recently observed in a discussion of judicial nominating commissions: A lack of transparency is highly damaging to the public's perception of the commission system. In the absence of information regarding proceedings, the public tends to think that the system is 'closed,' and that judges are selected through 'the old-boy system' or some other process that has little to do with the qualifications of the candidate. Such perceptions undermine the confidence in the quality of judges and ultimately in the quality of the legal system.

Third, public confidence is also promoted by the simple fact that most judges historically do well on evaluations, and the federal bench—given the increased salaries, prestige, and difficulty of appointment—should shine in this area. Certainly at the state level, judges selected through a competitive appointive process tend to do well on evaluations.¹⁹⁹ Federal district judges in particular, whose initial qualifications were sufficient to elicit a Presidential appointment and Senate confirmation, are presumed to be well-qualified to take the bench. Dissemination of evaluation results serves both to confirm this presumption, and to demonstrate that each judge is motivated to grow and improve on the bench.

Fourth, for those judges who do not receive a strong evaluation, public dissemination may serve as motivation to improve performance. Indeed, it may be the single strongest source of motivation. Federal district judges do not face the danger of being removed in an election; only personal standards of excellence and the risk of public embarrassment align a judge toward continued improvement on the bench. For most judges (state and federal), personal drive is enough. For those few judges who have lost their enthusiasm for the job, however, publication of poor evaluation results might provide the appropriate kick-start to either dedicate oneself to rapid professional improvement, or resign from the bench if that dedication is lacking.

Finally, widespread dissemination assures that the information the public actually receives on its judges is comprehensive and accurate. Public awareness of the federal judiciary has been influenced by the proliferation of media coverage for judicial nominations, confirmation hearings, and infamous cases.²⁰⁰ Furthermore, in recent years an increasing number of websites have emerged that either target specific judges²⁰¹ or ask for anonymous and unedited comments on a wide range of judges.²⁰²

Jeffrey D. Jackson, *Beyond Quality: First Principles in Judicial Selection and Their Application to a Commission-Based System*, 34 *FORDHAM URB. L.J.* 125, 157 (2007). The identical point applies to judicial evaluation commissions—full disclosure of the evaluation process and the identities of the commission members promotes public understanding of the system and confidence therein.

199. Alaska, for example, has utilized merit selection since statehood and has used a JPE program since 1975. During that time, only three judges have not been recommended for retention, in part because of the careful efforts of the nomination commission to recommend highly qualified candidates. Bill Gordon, Member, Alaska Judicial Council, Remarks at IAALS Symposium: Judicial Performance Evaluation: Strategies for Success (Aug. 5, 2008) (notes of remarks on file with authors).

200. This list of federal cases generating high media interest in the last several years alone is extensive and springs easily to mind, particularly in the areas of accused corporate malfeasance, criminal activity by celebrities, or terrorism. The judges presiding over these cases have found themselves under an unexpected (and uninvited) microscope, with greatly increased public scrutiny.

201. *E.g.*, Impeach Judge Turk!, <http://www.anusha.com/turk.htm> (last visited Oct. 17, 2008); Ten Federal Judges Who Must Be Impeached for Abuse of Power, <http://home.earthlink.net/~dlaw70/top10.htm> (last visited Oct. 17, 2008).

202. *See, e.g.*, The Robing Room, <http://www.therobingroom.com> (last visited Oct. 17, 2008). The website describes itself as “a site by lawyers for lawyers.” *Id.* Their mission is “to provide a forum for evaluating federal district court judges and magistrate-judges.” *Id.* Lawyers are encour-

The evaluations on these sites are at best unscientific and anecdotal, reflecting only the comments of a self-selected group; at worst, they are factually incorrect and motivated by personal bias rather than any dispassionate evaluation. JPE, by contrast, holds both the judges and the evaluation process *itself* to rigorous standards, creating a system in which both the public and the judiciary can be confident.

CONCLUSION

If thoughtfully implemented, a JPE program holds great promise both for the federal judiciary and for those who use and rely on the federal courts to preserve their rights and resolve their disputes. Pilot programs, drawn from elements of successful programs at the state level, may be useful in developing a program tailored to the unique needs of the federal system. A well-constructed JPE program would help locate judicial accountability where it belongs—in the process of adjudication—rather than in public or Congressional reaction to case outcomes.

Ultimately, for any JPE program to be accepted and functional, it must be embraced both by federal judges and the public they serve. Individual judges will need to step forward as leaders in this enterprise. The benefits of doing so—reframing judicial accountability, preserving judicial independence, and building public confidence in the courts—suggest that such an effort would be well worthwhile.

