

Judicial Performance Evaluation in Utah: Preliminary Findings*

Rebecca D. Gill
University of Nevada Las Vegas
rebecca.gill@unlv.edu

Rafael Oganessian
University of Nevada Las Vegas
oganesya@unlv.nevada.edu

March 28, 2015

Abstract

Judicial Performance Evaluation (JPE) is generally seen as an important part of the merit system, which often suffers from a lack of relevant voter information. Utah's JPE system has undergone significant change in recent years. Using data from the two most recent JPE surveys, we provide a preliminary look at the operation of this new system. Our results suggest that the survey component has difficulty distinguishing among the judges on the basis of relevant criteria. The question prompts intended to measure performance on different ABA Categories are also indistinguishable. We also find evidence that, on some measures, female judges do disproportionately worse than male judges. We suggest that the free response comments and the new Court Observation Program results may improve the ability of the commission to make meaningful distinctions among the judges on the basis of appropriate criteria.

1 Introduction

States have long been concerned with designing judicial selection systems that strike the right balance between independence and accountability. In the last several decades, many states have adopted a merit plan selection system (Gill 2013). This system pairs appointment by nonpartisan commission with retention by uncontested, nonpartisan election. A side effect of this arrangement is that voters are significantly less engaged in these noncompetitive elections (Bonneau and Hall 2009). This is due, in part, to the dearth of candidate information in noncompetitive elections (Klien and Baum 2001).

*Paper prepared for presentation at the 2015 Annual Meeting of the Western Political Science Association, Las Vegas NV. This research was supported by a grant from the National Science Foundation (SES 1354544). The authors also wish to thank the Judicial Performance Evaluation Commission and its Executive Director, Joanne C. Slotnik, for invaluable assistance in obtaining raw data.

In contested partisan elections, voters will minimize the cost of voting by relying on partisan cues (Downs 1957). In contested nonpartisan elections, voters rely upon a number of cues to serve as proxy measures for political party (Dubois 1984). Indeed, partisan considerations are an important part of nonpartisan elections (Hall 2001; Streb 2007). In both types of competitive elections, increased campaign spending is associated with higher voter participation, most likely because of the expensive campaign's ability to provide voters with information (Hall and Bonneau 2008).

Voters in retention elections generally have even less information available to them. Indeed, these voters will latch onto nearly any partisan cue in an attempt to formulate an opinion in these races (Squire and Smith 1988). In reaction to this problematic lack of information, many supporters of the merit system encourage the use of judicial performance evaluation systems (JPEs) (Brody 2000; Dubofsky 2007; White 2001). The hope is that properly administered JPEs can increase electoral awareness, depoliticize the selection process, and provide some measures of judicial accountability to the voters (Kourlis and Singer 2007; Paynter and Kearney 2010; White 2009). Indeed, some even recommend the use of JPEs in competitive judicial elections (Singer 2007).

Others have expressed deep concern about the use of JPEs as a part of the official process of judicial retention (Durham 2000; Griffin 1994; Gill and Retzl 2014). Despite the existence of guidelines for JPEs (American Bar Association 2005), recent inquiries into the administration of JPEs have found numerous problems with the design and implementation of JPEs (Elek, Rottman, and Cutler 2012; Gill 2014; Sterling 1993).

There have been few careful studies of state-sponsored JPEs have been conducted in only a couple of states (Gill and Retzl 2014; Sterling 1993). As a result, we have precious little information about how well these systems operate in practice. In this paper, we add to this small but growing body of knowledge. Using an original database of judicial performance evaluations in Utah (2012-2014), we investigate the presence of gender bias in respondent questionnaires. Our results suggest that the survey component has difficulty distinguishing among the judges on the basis of relevant criteria. That said, we do find evidence of female judges being ranked lower on various categorical evaluations such as communication skills, administrative capacity, and professionalism and temperament.

2 Reliability and Validity of JPE Measures

The diversity of judicial selection systems across the American states represents the variety of ways in which states attempt to balance the core values of judicial accountability and judicial

independence. While all American systems provide accountability at least for judicial misconduct, most of them also try to hold judges accountable for the quality of their work (Gill 2013). To do this, the relevant decision makers must have access to some information about the quality of judge’s work (Dubois 1980). In addition, such information can serve as important feedback to help judges improve their performance (Aynes 1981; Chauvin 1989).

Since 1985, the American Bar Association has provided a series of Guidelines (American Bar Association 2005) for evaluating judges. It set out a series of performance dimensions on which judges should be rated. Although they do not provide verbatim recommended questions, the Guidelines do recommend the use of behavior-based evaluation instruments to “generate more meaningful information about judicial behavior” (American Bar Association 2005: 13). This is where things seem to go wrong (Bernick and Pratto 1995; Elek, Rottman, and Cutler 2012; Gill 2014).

Most states with official JPE programs have realized this. As a result, there has been a movement to supplement the longstanding practice of polling the bar (Feeney 1987) by including multiple sources and types of information in the evaluation process (Mahoney 1989; Woolf and Yim 2011). Even still, the survey of local attorneys remains the centerpiece of the JPE systems (Brody 2000), especially as performance evaluation commissions increase the number of laypeople involved in conducting the evaluations (Olson and Batjer 1999).

3 Utah: A Case Study

Utah adopted a merit-based judicial selection system on July 1, 1985. This system uses a system of Judicial Nominating Commissions to assemble a pool of qualified candidates. Each judicial district has its own nominating commission comprised of attorneys and non-attorneys selected by the governor. After reviewing the records and interviewing prospective candidates, the commissions refer a list of candidates (five for district courts and seven for appellate courts) to the governor. The governor selects an individual from the list to stand for confirmation by the Utah Senate (of the Courts 2010).

Following the appointment, the judge must stand for a retention election. According to Art. 8 §9 of the Utah Constitution, a judge “shall be subject to an unopposed retention election at the first general election held more than three years after appointment.” If the judge is retained following the initial election, then the judge will stand for subsequent retention elections every six years (or, for supreme court justices, every ten years). However, prior to standing for these retention elections, the Utah Judicial Council (UJC) assesses the performance of judges and renders a

decision on the qualifications of each judge. The performance evaluation reports are summarized and distributed publicly.

The JPE program was first introduced in 1986, and the first evaluation cycle commenced in 1990. The purpose of the program lay primarily in providing reliable information for voters and providing a channel for self-improvement. Initially, the UJC was charged with administering the performance evaluation survey. At the time, the UJC was comprised of thirteen commissioners, twelve of whom were judges and one of whom was a member of the bar commission (Esterling and Sampson 1998). The centerpiece of these early evaluations was the attorney survey of judicial performance. The analysis of these surveys was conducted by consultants, and the UJC would meet to vote on retention recommendations in light, primarily, of the survey results.

This practice continued for almost twenty years. In 2008, the Utah legislature created the Judicial Performance Evaluations Commission (JPEC). The JPEC differed from its predecessor in that it included significant non-judge membership. The thirteen member JPEC now includes appointees by the Utah Supreme Court, the state legislature, and the governor. The JPEC now evaluates respondent surveys from attorneys, courts staff, and jurors. The JPEC has also implemented a ground-breaking courtroom observation program that focuses heavily on maintaining procedural fairness and objectivity (Woolf and Yim 2011). The JPEC uses a set of evaluation criteria that resemble somewhat the American Bar Association (American Bar Association 2005) guidelines. These are presented in Table 2

[Table 1 about here.]

4 Perception of JPEs in Utah

Despite the objective of enhancing voter and judicial decision-making, previous analysis of the voter and judicial perceptions of JPEs has found startling differences between these two groups. In their work analyzing JPEs across four states, Esterling and Sampson (1998) found a generally positive assessment of JPEs by voters in Salt Lake City. Around fifty percent of the respondents were familiar with UJC, and another forty percent had obtained a copy of the evaluation report it produced. Of those who had knowledge of the report, seventy-three percent said their vote function was either partially or fully influenced by the survey. This suggests that judicial retention evaluations are serving their purpose of decrease the cost of information. For decreasing the cost of voting, sixty-eight percent of respondents said that retention evaluations will increase their likelihood of voting in a judicial election.

In all, Esterling and Sampson (1998) Utah respondents were overwhelmingly positive in the receptive of judicial evaluations which according to them increases the quality of judicial candidates and accountability. The high rates of evaluation report knowledge and positive perception can be partially due to the multiple methods of retention information dissemination including mail delivery and newspaper print. Over seventy-one percent of respondents noted that they acquired the official evaluation report through either newspapers or at their place of residence.

While the administered survey found that the public had a positive impression of the evaluations, a survey of Utah judges suggested otherwise. Judges were found to be overwhelmingly skeptical of the process, despite acknowledging the benefit of marginal information that is created by JPEs. Their skepticism was oriented towards the intended effect of the performance evaluations. Only about a quarter of the judges surveyed reported that evaluations would increase voter turnout in retention elections, while 29% agreed that voters rely on JPEs towards their vote choice. Interestingly, these figures were the lowest across the four states in the Esterling and Sampson (1998) analysis.

The skepticism and negativity present in the Esterling and Sampson (1998) survey may be associated with the procedural mechanisms and fairness of the JPE process. When asking if judges have access to a fair appeals process if they disagreed with the report, only a third agreed. A common complaint is the fact that only negative evaluations could request a interview with the commission. In addition, half of the judges interviewed found the evaluation process to undermine their judicial independence. For Esterling and Sampson (1998), the responses from Utah's judges warranted additional investigation. In a follow-up interview, judges voiced their concern of the attorney monopoly present in JPEs at that time.

Judges saw the JPEs as a popularity contest with "nice" judges scoring well. Despite their concern of the overwhelming impact of attorneys, when asked whether these evaluations impacted their sentencing practices, only 5.9 percent answered yes. This suggests that judges, while critiquing the survey for attempting to influence judicial sentencing behavior, fail to change their sentencing practices despite acknowledging that sentencing practices impacts attorney-based JPEs.

The Esterling and Sampson (1998) report demonstrates a discrepancy between perception of voters and judges on the effectiveness of JPEs. The disaffection of Utah's judges played heavily in the state's revision of its JPE program in 2008. As Woolf and Yim (2011) describe it, the major revision included an overhaul of the JPEC procedures and the addition of the Courtroom Observation Program (COP). The COP provides a qualitative assessment of procedural fairness in each judge's courtroom. The COP involves the observation of judges by at least four different volunteers.

They observe in a single-blind manner, such that the judge is not notified of the presence of the observer in the courtroom. They observe and rate the judge on measures of procedural fairness, and they provide descriptive content analysis. The observer would analyze whether the judge provided visitors of the court with voice, neutrality, respect, and trust.

Following pilot tests of the observation program, twenty criteria points were created and divided between six groups to provide a framework for the observer's content analysis. These groups summarize six key points: behavior, tone, neutrality, voice, communication, and apparent fairness. The content analysis of observers is then translated via computer software in order to provide a general synopsis of the observations. The software analyzes the aggregate observations and comes up with widely agreed-upon themes, minority observations, and anomalous comments, all of which get recorded in the judges' profile (Woolf and Yim 2011).

This qualitative mechanism of judicial evaluation has partially resolved the criticism brought forth by judges who critiqued previous manners of JPEs for their overreliance on quantitative-driven assessment which were done in retrospect, thus lacking a real-time analysis of judicial courtroom behavior. Thus, a balance was reached by the committee between including quantitative surveys and qualitative observations as a better-rounded manner in judicial behavior observation.

5 Data and Methodology

Since 1990, Utah has administered JPE surveys every two years to a portion of its judges. Because each individual judge is evaluated only once every six (or ten) years, each year's administration evaluates approximately a third of the sitting judges in the state. After undergoing a major overhaul in 2008, the JPEC began conducting the attorney, staff, and juror surveys online. For this reason, the JPEC has collected and maintained electronic data only for the 2012 and 2014 administrations of the evaluations. These data are available across all levels of the Utah judiciary, including the Utah Supreme Court.

Utah's judiciary is mostly white and male. A total of twenty-four judges stood for retention in 2012. In 2014, forty-seven judges stood for retention. In all, sixteen of these judges (23%) were women. Only two judges during this period were categorized by the JPEC as non-white; both of these judges were of Asian descent. This comes as little surprise, given the findings of the most recent U.S. Census that Utah's population in 2013 was approximately 92% white (U.S. Census Bureau N.d.). A total of nine judges during these election cycles were considered part-time, and they were excluded from the JPE process.

5.1 Data Sources

The 2012-2014 Judicial Performance Evaluation survey results were obtained in response to a Utah Government Records Access and Management Act (GRAMA) request. Data on judicial demographics including gender, race, education, law school, and court type were obtained from the American Bench Database (*The American Bench* N.d.). These data were supplemented through official biographies on the Utah Courts website.¹

In addition to this, we investigated the disciplinary records of the judges in our sample using publicly available data from the Utah Judicial Conduct Committee. There were no disciplinary records for the judges in our sample during the time periods relevant to their evaluation by the JPEC. This does not seem to be the result of perfectly exemplary behavior on the part of Utah's judges. Instead, the problem is the short time span in our sample and the short time that has elapsed since then.

Recently, Utah has seen a few episodes of less-than-professional judicial behavior. In 2010, Lehi County Judge Garry Sampson removed a gun from his holster while on the bench. He pointed it in the direction of his bailiff. Despite protestation that this brandishing of a firearm was done in a "joking manner," Judge Sampson was reprimanded for his actions (Carlisle 2010). Also in 2010, Judge Ronald R. Hare was charged with exposing himself in a restroom at a local park. In 2011, the state supreme court censured the judge and he resigned shortly thereafter (Falk 2011). A case of sexual harassment is still ongoing involving a now retired Weber County justice court judge, Craig Storey. Storey is accused of making unwanted advances and authoring sexually explicit poems to his office manager (Gurrister 2015). The woman's claims were met with inaction by the Judicial Conduct Commission in 2010 (Gurrister 2015).

Clearly, the official judicial conduct records do not tell the full story about the scandals and criticism that befall Utah's judges. For this reason, we supplement this information with a content analysis of newspaper coverage of the judges in our sample. Media coverage about the judges was compiled using judge name searches of the major newspapers in the state.²

5.2 The Dependent Variables

We aim to measure the impact of judicial characteristics on the results of judicial performance evaluations in Utah. As such, our dependent variables are operationalized as responses to specific

¹These biographies can be found here: <http://www.utcourts.gov/judgesbios/>. Last accessed March 28, 2015.

²These newspapers include the Salt Lake City Tribune and the Deseret.

questions on the 2012 and 2014 JPEC surveys. In 2012, the survey consisted of fifty-eight questions; in 2014, these questions were pruned by almost half to a total of twenty-nine. The full text of these questions is provided in Appendix A. There, we also show how we have aggregated these questions into the categories defined by the American Bar Association (2005) Guidelines.

Previous research has made clear that problems common to JPE survey design can compromise the validity, reliability, and perceived fairness of the resulting evaluations (Elek, Rottman, and Cutler 2012; Gill 2014). We begin by looking for some of the warning signs in the Utah JPEC survey data. The inter-dimensional correlation matrix in Table 3 shows a high alpha level among the categories and high, significant correlations among all five of the dimensions of judicial performance. Table 4 shows that most of the bivariate correlations within the dimensions are *lower* than the inter-dimensional correlations. The alpha scores show a similar pattern. This shows that, in terms of measuring distinct concepts, the questions within each ABA category are not any more similar to each other than they are to the questions in other categories.

[Table 2 about here.]

[Table 3 about here.]

Overall, attorneys tended to rate judges very highly on all of the measures of judicial performance. Table 5 summarizes the performance ratings and the individual respondent level, as well as after these responses have been aggregated by judge. The range of scores is from one (inadequate) to five (outstanding).³ The JPEC lists different minimum requirements for 2012 and 2014. In 2012, the judges must score at least a 3.0 out of 5.0 on at least 80% of the individual questions. In 2014, the JPEC requires a 3.6 in each of the evaluation categories⁴ as the minimum for the presumption of retention. None of the judges fell below these standards in either year. Table 5 does not show this directly. However, the minimum scores in all of the aggregate data are above the center value on the scale (which is 3).

[Table 4 about here.]

The JPEC survey also includes a section that prompts respondents to select from a list of adjectives those that best described the judge. The list included adjectives with both positive and

³The JPEC Reports, in which the retention scores are summarized, do not indicate that the anchors were for the scores 2-4. They give only the first and last anchor, which are inadequate and outstanding, respectively. It is not clear whether additional anchors appeared on the actual survey. See Gill (2014) for a discussion of the importance of these anchors.

⁴Here, the JPEC refers to its own categories, which we summarize in Table 2

negative connotations. The adjectives are listed in Table 6, along with individual level and aggregate level descriptive statistics. Taken together with the scores from Table 5, this provides strong evidence of a halo effect. Respondents selected positive adjectives between 20-43% of the time, while they used negative adjectives just 1-5% of the time.

[Table 5 about here.]

5.3 The Judge-Level Control Variables

In order to assess the JPE's ability to distinguish among judges on the basis of their on-the-job performance, we need provide exogenous measures of judicial quality to see if they predict JPE scores. To do this, we include a number of observable proxy variables. Descriptive statistics for these variables are included in Table 7. We include a measure of the prestige of the judge's law school alma mater, which we operationalize as an ordinal variable based upon groupings the 2014 U.S. News and World Report rankings. The distribution of the ranking of the judges' alma maters is displayed in Figure 1.⁵ These measures are intended to serve as proxy variables for qualities like legal ability, communication skills, and administrative skills (?). We also include each judge's years of experience on the bench at the time of the evaluation.

[Figure 1 about here.]

We also include measures of the amount and nature of the media coverage of each of the judges. We derive these measures from the local newspapers. The measures are counts of the number of articles about the judge. Each article was subjected to a simple content analysis by human coders to determine the tone of the article. Several categories of content were derived. Critical articles are those that expressed dissatisfaction with how the judge discharged official duties related to judicial decision making. Our original coding scheme distinguished these from articles tying the judge to some sort of scandal, which would include accusations of misconduct or insinuations that the judge is involved in some sort of personal shenanigans. Positive articles are those that mention the judge in a positive light, be it for the decisions the judge made or the contributions the judge has made to society. We also collected the raw number of neutral mentions of the judge in order to control for the possibility that name recognition is enhanced by mention of the judge's name in print.⁶ Finally,

⁵We have coded these as follows: Top 14 = 1; 15-50 = 2; 51-100 = 3, 100-end of rankings = 4, unranked = 5, and no J.D. = 6. We have opted to follow the U.S. News and World Report convention and leave the most prestigious schools with low scores and the least prestigious schools with high scores.

⁶We have also collected data about official reprimands. On our sample, only one of the seventy-one judges had ever been reprimanded, and that reprimand happened years before the sample period.

we include an indicator for whether the judge was serving as a member of the state supreme court bench at the time of the evaluation. This allows us to control for differences in media coverage volume, visibility, and other unobserved sources of variation associated with occupying the highest bench in the state. Similarly, and because of the significant changes in the survey between 2012 and 2014, we include an indicator variable for the year of the survey.⁷

[Table 6 about here.]

We are particularly interested in the role that immutable characteristics play in shaping how judges are evaluated. To determine the judge’s gender, we relied upon the pronouns used in the official judge biographies. Our indicator variable is for female judges, such that they are coded as ‘1.’ We also coded a race variable, which was derived in part from the raw survey database we were provided by the JPEC. That dataset had an indicator for judges who were considered by the JPEC to be of a minority race or ethnic group. In all, there were only two judges in this category, and both were of Asian descent. We omit this variable from the analysis for lack of variation.

Unfortunately, the survey designers declined to collect important demographic data about the survey respondents. For this reason, we are unable to make any determinations about the interaction between respondent gender and judge gender in predicting outcomes. However, the designers did collect a few pieces of information about the respondents. All of these measures are relevant only to the attorney respondents. They include the number of years in legal practice (grouped into six categories) and the number of trials the attorney has argued (also grouped into six categories). It also includes dummy variables to indicate the fields of law in which the respondent attorney specializes in. These fields include collections, family law, criminal law, civil trials, or other. We also include a dummy variable for whether the respondent was an attorney or a layperson. These variables are summarized in 8.

[Table 7 about here.]

6 Analyses of JPE Outcomes

We begin by investigating which criteria respondents use to distinguish among the judges. We have assembled a number of different measures that we hypothesize will be related to the judge evaluation scores by category. We construct a series of multivariate models with the scores on the various ABA Guidelines categories as the dependent variables. These models are summarized in

⁷Because the shortest possible time span between retention elections is six years, we have no judges in the sample who appear in more than one year’s survey.

Table 9. The models perform relatively poorly. They explain only a tiny sliver of the variance in the dependent variables.

[Table 8 about here.]

A few patterns do emerge. Figure 2 presents a summary of the coefficients for the models of the scores by ABA Guidelines (American Bar Association 2005) category. The scores seem to depend heavily upon the characteristics of the respondent. Recall that the performance is evaluated on a five-point scale. Overall, lawyers rate judges more harshly than laypeople. This is especially true when it comes to legal ability, where lawyers rate judges nearly three-quarters of a point lower. Lawyers who practice in criminal law even harsher than their peers, especially in their ratings on the integrity and impartiality scale.

The prestige of the judge’s law school—one of the stronger proxy variables of performance on several of the ABA categories—is completely unrelated to evaluation scores. More time on the bench does not appear to help, except when it comes to ratings of professionalism and temperament. The most damaging variable is the scandal variable. There was only one reported scandal in the dataset. It involved Judge Kevin Allen of the first district court. During a committee hearing, Judge Allen’s financial obligations were brought to light including several legal decisions against him. In addition, Judge Allen’s real estate dealings were also questioned including his recent rezoning dispute over a farmland in Smithfield City. Despite voicing skepticism on how Judge Allen can meet his financial obligations with a judges salary, the committee ultimately confirmed Judge Allen.

Although the magnitude is relatively small, female judges are rated lower on communication skills, administrative capacity, and professionalism/temperament. This is in keeping with previous research (Burger 2007; ?; Gill and Retzl 2013).

Female judges are generally rated lower on the measures of communication skills, administrative capacity, and professionalism/temperament.

[Figure 2 about here.]

As social cognition theory would predict, the assignment of some of these descriptive labels varies by the gender of the judge. Figure 3 and Figure 4 show the usage rates broken down by gender. Male judges are significantly more likely to be described as knowledgeable and intelligent, while female judges are significantly more likely to be described as attentive, patient, and receptive. On the negative side, men are significantly more likely to be described as dismissive and arrogant. The differences in the use of negative adjectives are relatively small in magnitude, of course, because

so few respondents chose to use these adjectives to describe judges.

7 Discussion

In all, the JPEC's judicial performance evaluation survey results in very little distinction among the judges. The halo effect is pronounced on the measures intended to capture the performance outlined in the ABA Guidelines. The most obvious covariates of judicial behavior explained very little of the variation in the ratings, perhaps because there was so little variation to begin with. Only 8% of respondents indicated that they felt the judge should not be retained. Despite this fact, the gender disparity survives in some of the measures of judicial performance.

It is also possible that more information is hidden in the other parts of the survey. For example, the adjective prompts yielded some interesting differences between the evaluation of female and male judges. In future research, we hope to explore the content of the free response and adjective questions. In addition, Utah's newly-implemented courtroom observation program provides rich qualitative data about how judges are perceived by trained, neutral observers (Woolf and Yim 2011).

When it comes down to it, however, the JPEC evaluations from 2012-2014 recommended for retention every single judge it evaluated. Perhaps this is because the judges are of uniformly high quality. Indeed, performance evaluation systems like this avoid erring on the side of harshness in recommendations. This is likely an attempt to preserve the independence of the judiciary and to encourage cooperation from the judges being evaluated. As well, if the proper role of the JPE is to act as a warning alarm for particularly low-performing judges, the exercise may not be in vain. However, this JPE survey yields precious little actionable feedback by distinguishing among the judges on the ABA evaluative categories.

[Figure 3 about here.]

[Figure 4 about here.]

References

- American Bar Association. 2005. Guidelines for the Evaluation of Judicial Performance with Commentary. Technical report ABA Chicago, IL: .
- Aynes, Richard L. 1981. "Evaluation of Judicial Performance: A Tool for Self-Improvement." *Pepperdine Law Review* 8(2): 255–312.
- Bernick, E. Lee, and David J. Pratto. 1995. "A Behavior-based Evaluation Instrument for Judges." *Justice System Journal* 18(2): 173–184.
- Bonneau, Chris W., and Melinda Gann Hall. 2009. *In Defense of Judicial Elections*. Controversies in Electoral Democracy and Representation New York: Routledge.
- Brody, David C. 2000. "Judicial Performance Evaluations by State Governments: Informing the Public While Avoiding the Pitfalls." *Justice System Journal* 21: 333–356.
- Burger, Gary K. 2007. "Attorney's Ratings of Judges: 1998-2006." Report to the Mound City Bar.
- Carlisle, Nate. 2010. "Lehi Judge Pulls Gun on Bailiff in Court." *Salt Lake City Tribune* (February 24).
- Chauvin, L. Stanley. 1989. "Judicial-Performance Evaluation." *American Bar Association Journal* 75(December): 8.
- Downs, Anthony. 1957. *An Economic Theory of Democracy*. New York: Harper.
- Dubofsky, Jean. 2007. "Judicial Performance Review—A Balance Between Judicial Independence and Public Accountability." *Fordham Urban Law Journal* 2007(January): 321–339.
- Dubois, Phil. 1980. *From Ballot to Bench: Judicial Elections and the Quest for Accountability*. Austin, TX: University of Texas Press.
- Dubois, Philip L. 1984. "Voting Cues in Nonpartisan Trial Court Elections: A Multivariate Assessment." *Law and Society Review* 18(3): 395–436.
- Durham, Christine M. 2000. "Gender and Professional Identity: Unexplored Issues in Judicial Performance Evaluation." *Judges' Journal* 39(2): 13–16.
- Elek, Jennifer K., David B. Rottman, and Brian L. Cutler. 2012. "Judicial Performance Evaluation: Steps to Improve Survey Process and Measurement." *Judicature* 95: 65–75.
- Esterling, Kevin M., and Kathleen M. Sampson. 1998. *Judicial Retention Evaluation Programs in Four States: A Report with Recommendations*. Chicago, IL: American Judicature Society.

- Falk, Aaron. 2011. "Utah Supreme Court Sanctions Two Judges in FY 2011." *Salt Lake City Tribune* (October 3).
- Feeney, Floyd. 1987. "Evaluating Trial Court Performance." *Justice System Journal* 12(1): 148–170.
- Gill, Rebecca. 2014. "Implicit Bias in Judicial Performance Evaluations: We Must Do Better than This." *Justice System Journal* Online: 1–24.
- Gill, Rebecca D. 2013. "Beyond High Hopes and nmet Expectations: Judicial Selection Reforms in the States." *Judicature* 96(6): 278–295.
- Gill, Rebecca D., and Kenneth J. Retzl. 2013. "Implicit Gender Bias in State-Sponsored Judicial Performance Evaluations: A Preliminary Analysis of Colorado's JPE System, 2002-2012." .
- Gill, Rebecca D., and Kenneth J. Retzl. 2014. "The JPE Commission: Toward a More Transparent and Informative Evaluation." *Judicature* 98(1): 26–34.
- Griffin, Jacqueline R. 1994. "Judging the Judges." *Litigation* 21(3): 5.
- Gurrister, Tim. 2015. "Woman Suing Weber County Claims More Retaliation Against Her." *Standard Examiner* (March 24).
- Hall, Melinda Gann. 2001. "State Supreme Courts in American Democracy: Probing the Myths of Judicial Reform." *American Political Science Review* 95(2): 315–331.
- Hall, Melinda Gann, and Chris W. Bonneau. 2008. "Mobilizing Interest: The Effects of Mondy on Citizen Participation in State Supreme Court Elections." *American Journal of Po* 52(July): 457–470.
- Klien, David, and Lawrence Baum. 2001. "Ballot Information and Voting Decisions in Judicial Elections." *Political Research Quarterly* 54: 709–728.
- Kourlis, Rebecca Love, and Jordan M. Singer. 2007. "Using Judicial Performance Evaluations to Promot Judicial Accountability." *Judicature* 90(5): 200–207.
- Mahoney, Anne Rankin. 1989. "Citizen Evaluation of Judicial Performance: The Colorado Experience." *Judicature* 72(4): 210–216.
- of the Courts, Administrative Office. 2010. "Manual of Procedures for Justice Court Nominating Commissions." .
- Olson, Susan M., and Christina Batjer. 1999. "Competing Narratives in a Judicial Retention Election: Feminism versus Judicial Independence." *Law & Society Review* 33(1): 123–160.

- Paynter, Sharon, and Richard C. Kearney. 2010. "Who Watches the Watchmen? Evaluating Judicial Performance in the American States." *Administration and Society* 41(8): 923–953.
- Singer, Jordan M. 2007. "Knowing is Half the Battle: A Proposal for Prospective Performance Evaluations in Judicial Elections." *University of Arkansas Little Rock Law Review* 29: 725.
- Squire, Peverill, and Eric R.A.N. Smith. 1988. "The Effect of Partisan Information on Voters in Nonpartisan Elections." *Journal of Politics* 50(1): 169–179.
- Sterling, Joyce S. 1993. "The Impact of Gender Bias on Judging: Survey of Attitudes Toward Women Judges." *Colorado Law Review* 22: 257.
- Streb, Matthew J. 2007. "Partisan Involvement in Partisan and Nonpartisan Trial Court Elections." In *Running for Judge: The Rising Political, Financial, and Legal Stakes of Judicial Elections*, ed. Matthew J. Streb. New York: New York University Press chapter 6.
- The American Bench*. N.d. Online version.
- U.S. Census Bureau. N.d. "State and County QuickFacts." Online.
- White, Penny J. 2001. "Judging Judges: Securing Judicial Independence by Use of Judicial Performance Evaluations." *Fordham Law Review* 29: 1053.
- White, Penny J. 2009. "Retention Elections in a Merit-Selection System: Balancing the Will of the Public with the Need for Judicial Independence and Accountability: Using Judicial Performance Evaluations to Supplement Inappropriate Voter Cues and Enhance Judicial Legitimacy." *Missouri Law Review* 74: 635.
- Woolf, Nicholas H., and Jennifer M.J. Yim. 2011. "The Courtroom-Observation Program of the Utah Judicial Performance Evaluation Commission." *Court Review* 47: 84–91.

Appendix A JPEC Questions and ABA Categories

Survey Questions	ABA Category
2012	
The Judge makes sound rulings.	5-1: Legal Ability
The judge properly applies the rules of civil procedure.	5-1: Legal Ability
The judge properly applies the rules of criminal procedure.	5-1: Legal Ability
The judge properly applies the rules of evidence.	5-1: Legal Ability
The judge's sentencing fits the offenses.	5-1: Legal Ability
The judge makes appropriate findings of facts.	5-1: Legal Ability
The judge appropriately applies the laws to the facts.	5-1: Legal Ability
The judge follows legal precedent.	5-1: Legal Ability
The judge only considers evidence in the record.	5-1: Legal Ability
The judge's written decisions are clear and logical.	5-3: Communication Skills
The judge's written opinions offer meaningful legal analysis.	5-1: Legal Ability
The judge was fair and impartial.	5-2: Integrity and Impartiality
The judge avoids impropriety and the appearance of impropriety.	5-2: Integrity and Impartiality
The judge avoids improper ex parte communications.	5-2: Integrity and Impartiality
The judge's behavior demonstrated equal treatment of all persons or classes of persons.	5-2: Integrity and Impartiality
The judge appears to consider both sides of an argument before rendering a decision.	5-2: Integrity and Impartiality
The judge holds attorneys accountable for inappropriate conduct.	5-5: Administrative Capacity
The judge's oral communication while in court is clear and logical.	5-3: Communication Skills
The judge promotes public trust and confidence in the courts through his or her conduct on the bench.	5-4: Professionalism and Temperament
The judge respects the time of the participants and understands the personal and financial costs they may be incurring.	5-4: Professionalism and Temperament
The judge is prepared for argument and hearings.	5-5: Administrative Capacity
The judge treats all attorneys with equal courtesy and respect.	5-2: Integrity and Impartiality
The judge rules in a timely manner.	5-5: Administrative Capacity
Continued on next page	

bfseries Table 1 – continued from previous page

Survey Questions	ABA Category
The judge realistically manages his or her calendar.	5-5: Administrative Capacity
The judge convened court without undue delay.	5-5: Administrative Capacity
The judge provides the parties due process; namely, advance notice of issues to be heard an adequate opportunity to prepare and a meaningful opportunity to be heard.	5-4: Professionalism and Temperament
The judge acts to ensure that linguistic/cultural differences or disabilities do not unfairly limit access to the justice system.	5-5: Administrative Capacity
The judge is willing to make difficult or unpopular decisions.	5-2: Integrity and Impartiality
The judge did not allow his or her personal beliefs to inappropriately influence the proceedings.	5-2: Integrity and Impartiality
The judge explains the reasons for his or her decisions, when appropriate.	5-3: Communication Skills
The judge works with pro se litigants fairly and effectively.	5-4: Professionalism and Temperament
The judge’s personal life does not impair his or her judicial performance.	5-2: Integrity and Impartiality
The judge maintains diligent work habits.	5-5: Administrative Capacity
The judge’s interactions with court staff are professional and constructive.	5-4: Professionalism and Temperament
The judge is an effective manager of his or her staff, operations and business.	5-5: Administrative Capacity
The judge appropriately enforces deadlines and court orders.	5-5: Administrative Capacity
The judge is appropriately accessible to court personnel.	5-5: Administrative Capacity
The judge made sure that everyone’s behavior in the courtroom was proper.	5-5: Administrative Capacity
The judge reasonably accommodates changing technology.	5-5: Administrative Capacity
The judge paid attention to the proceedings in the courtroom.	5-5: Administrative Capacity
The judge is collegial with other members of the court.	5-4: Professionalism and Temperament
When the judge explained to the jury the reasons for his or her decision, I understood.	5-3: Communication Skills
Based on the judge’s explanations, I clearly understood my role and responsibility as a juror.	5-3: Communication Skills

Continued on next page

bfseries Table 1 – continued from previous page

Survey Questions	ABA Category
The jury instructions from the judge were clear and understandable.	5-3: Communication Skills
Based on the judge’s explanations, I understood the evidence I could or could not consider.	5-3: Communication Skills
The judge demonstrated courtesy toward the attorneys, court staff, litigants and others in the court room.	5-4: Professionalism and Temperament
The judge made me feel that the court system is fair.	5-2: Integrity and Impartiality
The judge took the case seriously.	5-4: Professionalism and Temperament
The judge treated the jury with respect.	5-4: Professionalism and Temperament
The judge provided recesses (breaks) in the trial that were adequate	5-5: Administrative Capacity
My experience with the judge helped me understand the role of the jury in the legal system.	5-4: Professionalism and Temperament
2014	
The judge follows the applicable legal rules (e.g. civil procedure, criminal procedure, evidence, juvenile, appellate) that apply to the case.	5-1: Legal Ability
The judge makes appropriate findings of fact and applies the law to those facts.	5-1: Legal Ability
The judge follows legal precedent or clearly explains departures from precedent	5-1: Legal Ability
The judge only considers evidence in the record.	5-1: Legal Ability
The judge’s written opinions/decisions offer meaningful legal analysis.	5-1: Legal Ability
The judge’s written opinions contain a fair statement of the pertinent facts.	5-2: Integrity and Impartiality
The judge’s written opinions contain a discussion of the applicable legal principles and controlling case law.	5-1: Legal Ability
The judge’s written opinions clearly address the merits of the legal issues advanced by the parties.	5-3: Communication Skills
The judge’s written opinions provide clear guidance to trial court judges and practitioners.	5-3: Communication Skills
The judge’s written opinions contain a readily understandable, concise ruling.	5-3: Communication Skills
Continued on next page	

bfseries Table 1 – continued from previous page

Survey Questions	ABA Category
The judge’s written opinions reflect a neutral, professional tone.	5-2: Integrity and Impartiality
The judge makes sure that everyone’s behavior in the courtroom is proper.	5-5: Administrative Capacity
The judge appears to pay attention to what goes on in court.	5-4: Professionalism and Temperament
The judge’s personal life or beliefs do not impair his or her judicial performance.	5-2: Integrity and Impartiality
The judge demonstrates respect for the time and expense of those attending court.	5-4: Professionalism and Temperament
The judge promotes access to the justice system for people who speak a language other than English, or for people who have a physical or mental limitation.	5-5: Administrative Capacity
The judge is prepared for court proceedings.	5-5: Administrative Capacity
The judge’s interactions with courtroom participants and staff are professional and constructive.	5-4: Professionalism and Temperament
The judge is an effective manager.	5-5: Administrative Capacity
The judge convenes court without undue delay.	5-5: Administrative Capacity
The judge rules in a timely fashion.	5-5: Administrative Capacity
The judge accommodates changing technology.	5-5: Administrative Capacity
The judge maintains diligent work habits.	5-5: Administrative Capacity
The judge’s oral communications are clear.	5-3: Communication Skills
The judge’s written opinions/decisions are clear and logical.	5-3: Communication Skills
The judge treats all courtroom participants with equal respect.	5-2: Integrity and Impartiality
The judge promotes public trust and confidence in the courts through his or her conduct.	5-4: Professionalism and Temperament
Continued on next page	

bfseries Table 1 – continued from previous page

Survey Questions	ABA Category
The judge provides the parties with a meaningful opportunity to be heard.	5-4: Professionalism and Temperament

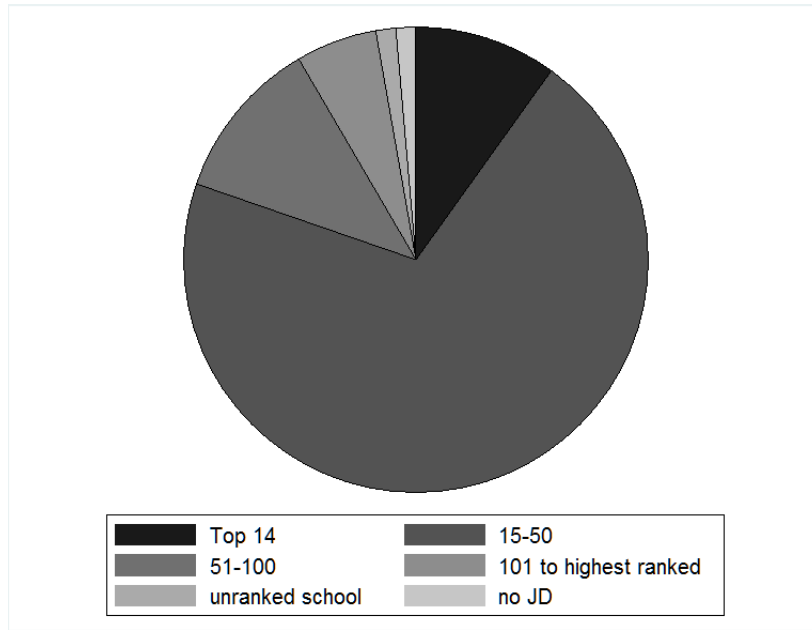


Figure 1: Law School Rank of Utah Judges Evaluated 2012-2014

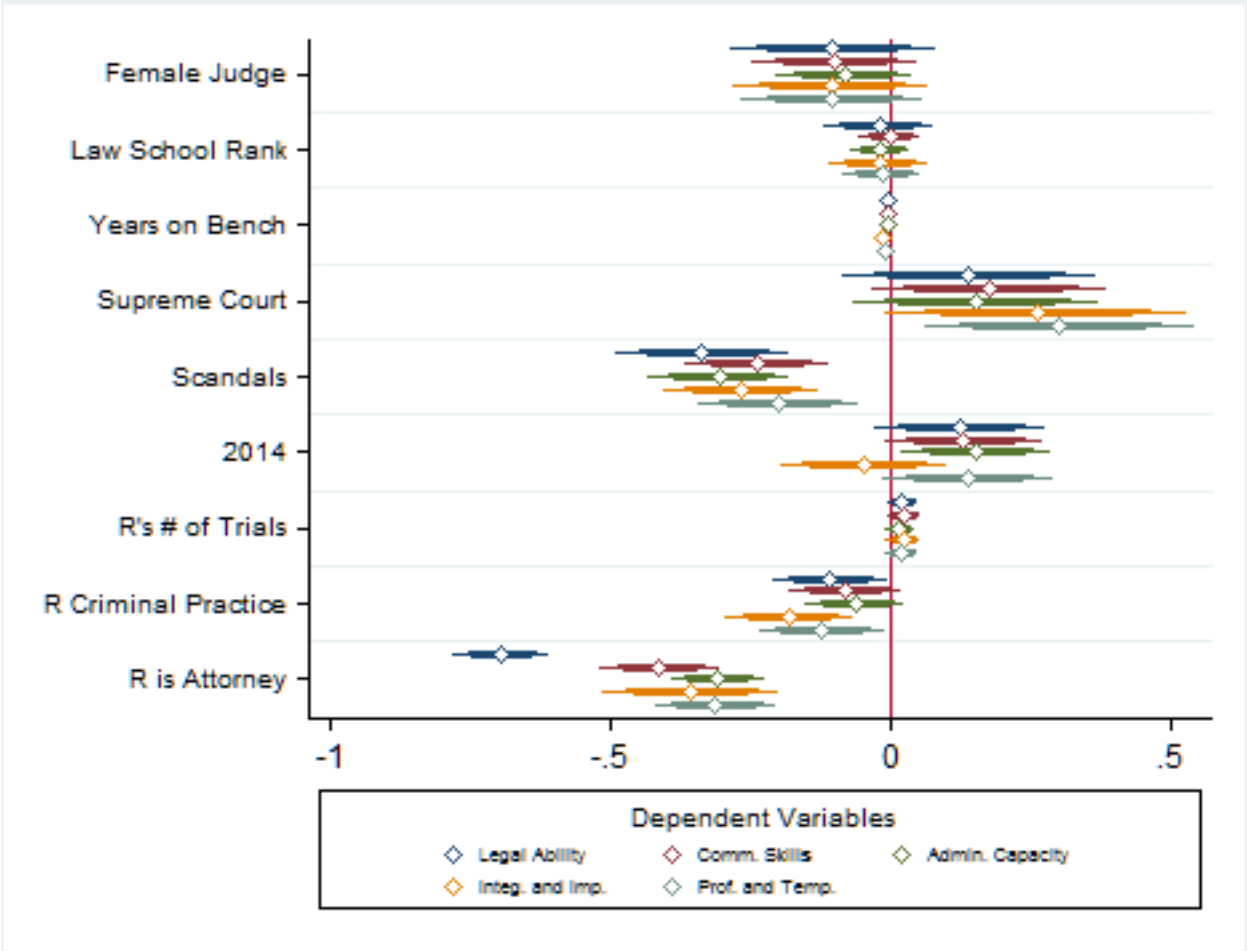


Figure 2: ABA Category Models: Coefficient Summaries

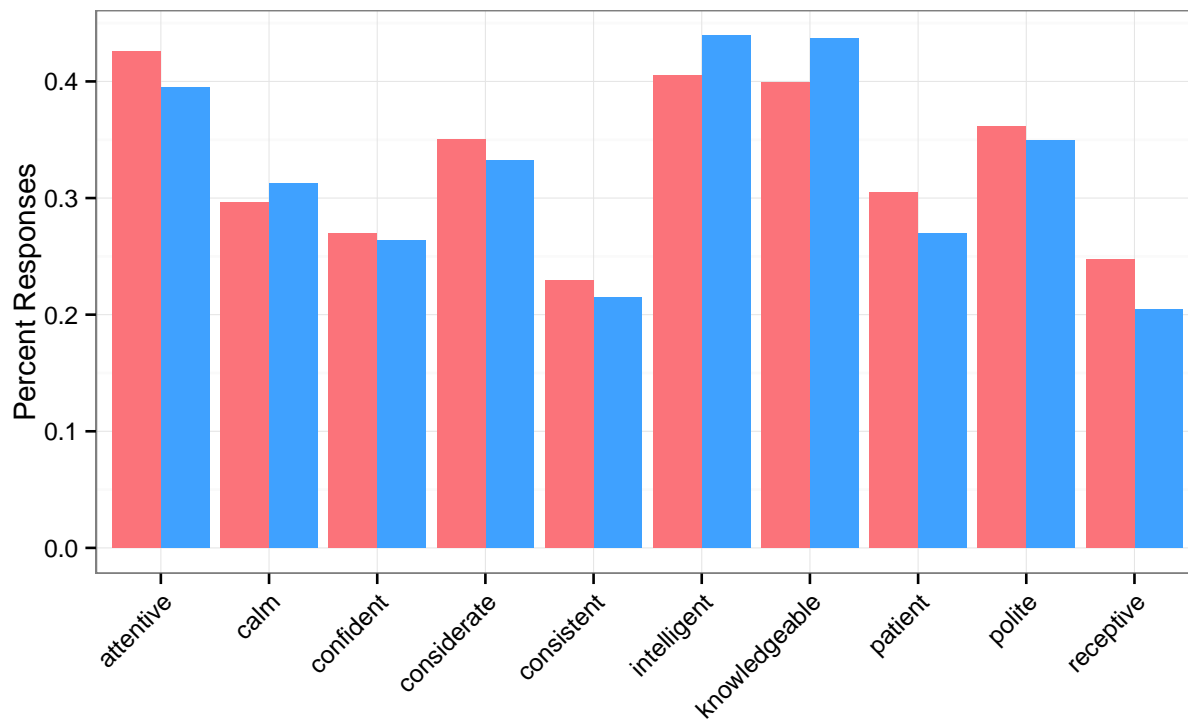


Figure 3: Use of Positive Adjectives by Judge Gender

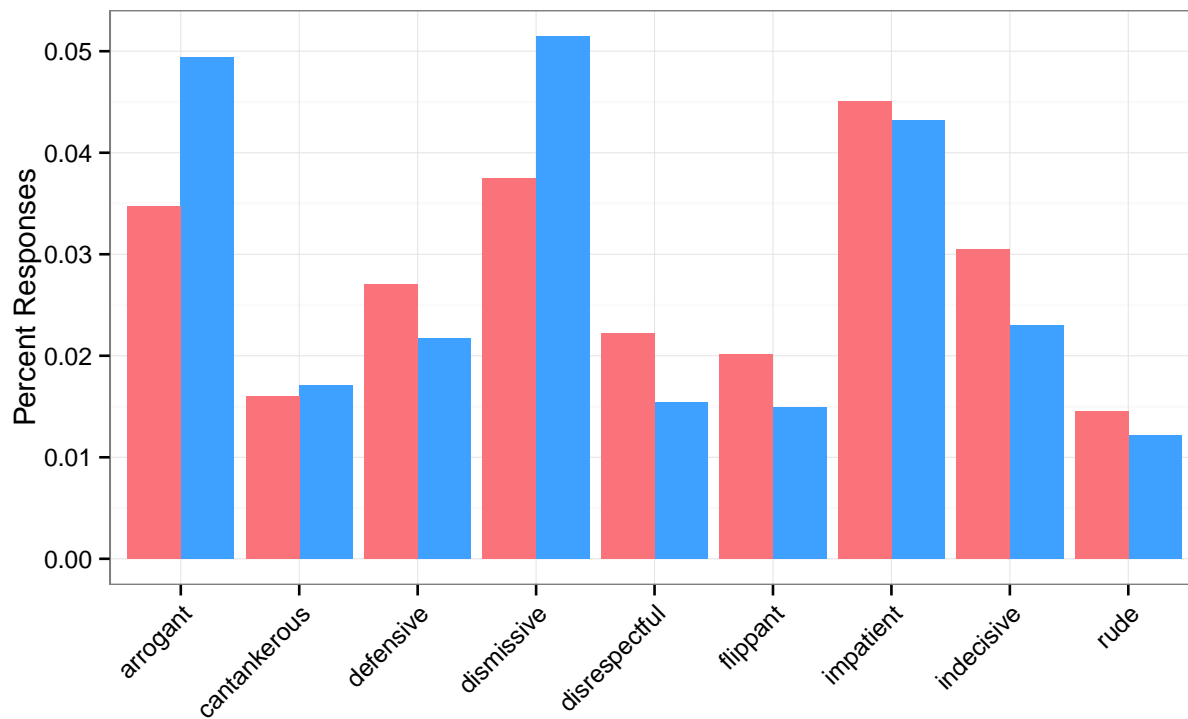


Figure 4: Use of Negative Adjectives by Judge Gender

Table 2: Criteria for JPEC Performance Evaluation

<p>Integrity</p> <ul style="list-style-type: none"> a) avoidance of impropriety and appearance of impropriety; b) freedom from personal bias; c) ability to decide issues based on the law and facts without regard to the identity of the parties or counsel, the popularity of the decision, and without concern for criticism; d) impartiality of actions; e) compliance with the Code of Judicial Conduct. 	<p>Knowledge and Understanding of Law</p> <ul style="list-style-type: none"> a) the issuance of legally sound decisions; b) understanding of the substantive, procedural, and evidentiary law of the state; c) attentiveness to the factual and legal issues before the court; d) the proper application of judicial precedents and other appropriate sources of authority.
<p>Ability to Communicate</p> <ul style="list-style-type: none"> a) clarity of bench rulings and other oral communications; b) quality of written opinions with specific focus on clarity and logic, and the ability to explain clearly the facts of a case and the legal precedents at issue; c) sensitivity to impact of demeanor and other nonverbal communications. 	<p>Preparation, Attentiveness, Dignity, and Control over Proceedings</p> <ul style="list-style-type: none"> a) courtesy to all parties and participants; b) willingness to permit every person legally interested in a proceeding to be heard, unless precluded by law.
<p>Skills as a Manager</p> <ul style="list-style-type: none"> a) devoting appropriate time to all pending matters; b) discharging administrative responsibilities diligently; c) sensitivity to impact of demeanor and other nonverbal communications. 	<p>Punctuality</p> <ul style="list-style-type: none"> a) the prompt disposition of pending matters; b) meeting commitments on time and according to the rules of the court; c) compliance with the case processing time standard established by the Council.
<p>Service to the Profession and the Public</p> <ul style="list-style-type: none"> a) attendance at and participation in judicial and continuing legal education programs; b) consistent with the Code of Judicial Conduct, participation in organizations devoted to improving the justice system; c) consistent with the highest principles of the law, ensuring that the court is serving the public and the justice system to the best of its ability and in such a manner as to instill confidence in the court system; d) service within the organizations of the judicial branch of government and in leadership positions within the judicial branch of government, such as presiding judge, Judicial Council, Boards of Judges, and standing and ad hoc committees. 	<p>Effectiveness in Working with Judges, Commissioners, Personnel</p> <ul style="list-style-type: none"> a) when part of a multi-judge panel, exchanging ideas and opinions with other judges during the decision-making process; b) critiquing the work of colleagues; c) facilitating the administrative responsibilities of other judges and commissioners; d) effectively working with court staff.

Table 3: Inter-Dimensional Correlation Matrix

Category	<i>Legal Ability</i>	<i>Integrity & Impartiality</i>	<i>Communication Skills)</i>	<i>Professionalism & Temperament</i>	<i>Administrative Capacity</i>
<i>Legal Ability</i>	1.000				
<i>Integrity & Impartiality</i>	0.846***	1.000			
<i>Communication Skills</i>	0.880***	0.809***	1.000		
<i>Professionalism & Temperament</i>	0.839***	0.902***	0.832***	1.000	
<i>Administrative Capacity</i>	0.807***	0.801***	0.822***	0.878***	1.000

Average inter-item covariance = 0.539, $\alpha = 0.967$

Table 4: Intra-Dimensional Analysis Summary Table

Category	<i>No. Questions</i>	<i>N Range</i>	<i>Avg. Cov.</i>	<i>Alpha</i>	<i>Corr. Range</i>
2012 Questions					
<i>Legal Ability</i>	10	781-1505	0.765	0.979	0.696-0.922
<i>Integrity & Impartiality</i>	6	1383-1963	0.616	0.956	0.676-0.863
<i>Communication Skills</i>	2	1148-1734	0.741	0.935	0.831
<i>Professionalism & Temperament</i>	3	1483-1752	0.681	0.985	0.698-0.775
<i>Administrative Capacity</i>	5	1062-1973	0.561	0.920	0.602-0.755
2014 Questions					
<i>Legal Ability</i>	6	548-2818	0.796	0.974	0.789-0.897
<i>Integrity & Impartiality</i>	4	540-3835	0.738	0.954	0.685-0.842
<i>Communication Skills</i>	5	545-3856	0.511	0.941	0.553-0.914
<i>Professionalism & Temperament</i>	5	2780-3821	0.528	0.943	0.681-0.836
<i>Administrative Capacity</i>	7	3213-5092	0.370	0.931	0.551-0.792

Full intra-dimensional correlation tables available from authors by request. Some questions listed in Appendix A are asked only of state supreme court judges. These questions yield 500 or fewer responses, and are omitted from this summary table.

Table 5: Responses over ABA Categories

Variable	Mean	Std. Dev.	Min.	Max.	N
<i>Individual Level</i>					
Legal Ability	4.208	0.904	1	5	4354
Comm. Skills	4.445	0.801	1	5	5887
Admin. Capacity	4.470	0.685	1	5	5914
Integ. & Imp.	4.460	0.820	1	5	5906
Pro. & Temp.	4.477	0.774	1	5	5914
<i>Aggregate Level 2012</i>					
Legal Ability	4.087	0.199	3.668	4.394	24
Comm. Skills	4.330	0.170	3.909	4.559	24
Admin. Capacity	4.338	0.185	3.923	4.618	24
Integ. & Imp.	4.420	0.178	3.995	4.708	24
Pro. & Temp.	4.345	0.202	3.807	4.685	24
<i>Aggregate Level 2014</i>					
Legal Ability	4.266	0.248	3.468	4.712	47
Comm. Skills	4.496	0.196	3.994	4.830	47
Admin. Capacity	4.528	0.160	4.051	4.784	47
Integ. & Imp.	4.458	0.254	3.746	4.808	47
Pro. & Temp.	4.528	0.210	3.955	4.823	47

Table 6: Adjective Prompts

Variable	Mean	Std. Dev.	Min.	Max.	N
<i>Individual Level: Positive Adjectives</i>					
knowledgeable	0.429	0.495	0	1	6881
intelligent	0.432	0.495	0	1	6881
attentive	0.402	0.490	0	1	6881
considerate	0.336	0.472	0	1	6881
calm	0.309	0.462	0	1	6881
confident	0.265	0.441	0	1	6881
patient	0.277	0.448	0	1	6881
consistent	0.218	0.413	0	1	6881
polite	0.352	0.478	0	1	6881
receptive	0.213	0.410	0	1	6881
<i>Individual Level: Negative Adjectives</i>					
impatient	0.044	0.204	0	1	6881
defensive	0.023	0.149	0	1	6881
rude	0.013	0.112	0	1	6881
cantankerous	0.017	0.129	0	1	6881
indecisive	0.025	0.155	0	1	6881
dismissive	0.045	0.208	0	1	6881
arrogant	0.046	0.210	0	1	6881
disrespectful	0.017	0.129	0	1	6881
flippant	0.016	0.125	0	1	6881
<i>Aggregate: Positive Adjectives</i>					
knowledgeable	0.436	0.140	0.157	0.814	71
intelligent	0.433	0.132	0.184	0.770	71
attentive	0.410	0.107	0.143	0.617	71
considerate	0.343	0.129	0.093	0.611	71
calm	0.316	0.116	0.111	0.583	71
confident	0.271	0.109	0.070	0.552	71
patient	0.283	0.131	0.019	0.684	71
consistent	0.231	0.104	0.060	0.583	71
polite	0.361	0.137	0.102	0.722	71
receptive	0.220	0.083	0.056	0.472	71
<i>Aggregate: Negative Adjectives</i>					
impatient	0.046	0.052	0	0.191	71
defensive	0.024	0.023	0	0.082	71
rude	0.013	0.022	0	0.102	71
cantankerous	0.019	0.033	0	0.163	71
indecisive	0.025	0.030	0	0.149	71
dismissive	0.044	0.042	0	0.185	71
arrogant	0.048	0.068	0	0.337	71
disrespectful	0.019	0.028	0	0.163	71
flippant	0.016	0.022	0	0.09	71

Table 7: Independent Variables

Variable	Mean	Std. Dev.	Min.	Max.	N
<i>Individual Level Data</i>					
female	0.209	0.407	0	1	6881
minority	0.019	0.137	0	1	6724
law school rank	2.129	0.759	1	6	6881
years on bench	9.874	7.096	2	32	6881
supreme court	0.03	0.172	0	1	6881
Media Coverage:					
scandal	0.013	0.113	0	1	6881
critical	0.2	0.507	0	3	6881
positive	0.466	0.739	0	3	6881
neutral	21.007	19.996	0	86	6881
year	2012	n=2883	2014	n=3998	N=6881
<i>Aggregate Data</i>					
female	0.225	0.421	0	1	71
minority	0.029	0.168	0	1	70
law school rank	2.225	0.865	1	6	71
years on bench	9.745	7.052	2	32	71
supreme court	0.028	0.167	0	1	71
Media Coverage:					
scandal	0.014	0.119	0	1	71
critical	0.197	0.521	0	2.958	71
positive	0.479	0.753	0	3	71
neutral	18.577	18.696	0	86	71
year	2012	n=24	2014	n=47	N=71

Table 8: Respondent Characteristics

Variable	Mean	Std. Dev.	Min.	Max.	N
Years Exp.	5.036	2.970	0	8	5508
No. Trials	1.555	1.488	0	5	5507
Collections	0.034	0.182	0	1	6282
Domestic	0.182	0.386	0	1	6282
Criminal	0.274	0.446	0	1	6282
Civil	0.380	0.486	0	1	6282
Other Area	0.061	0.239	0	1	6282
Attorney	0.751	0.432	0	1	6881

Table 9: Models of Judicial Performance on JPEC Survey by ABA Category: OLS with Huber-White Standard Errors

Variable	<i>Legal Ability</i>		<i>Communication Skills</i>		<i>Administrative Capacity</i>		<i>Integrity and Impartiality</i>		<i>F</i>
Female Judge	-0.106	(0.070)	-0.114*	(0.056)	-0.102*	(0.048)	-0.120	(0.068)	-0.120
Law School Rank	-0.019	(0.036)	-0.004	(0.021)	-0.024	(0.020)	-0.027	(0.032)	-0.027
Years on Bench	-0.004	(0.004)	-0.005	(0.003)	-0.003	(0.002)	-0.011**	(0.011)	-0.011**
Supreme Court Judge	0.145	(0.086)	0.167*	(0.076)	0.146	(0.083)	0.229**	(0.052)	0.229**
Scandal in News	-0.315***	(0.061)	-0.207**	(0.045)	-0.305***	(0.043)	-0.228***	(0.051)	-0.228***
2014 Survey	0.128*	(0.058)	0.132**	(0.044)	0.144***	(0.043)	-0.016	(0.052)	-0.016
R No. of Trials	0.021	(0.010)	0.026*	(0.011)	0.017	(0.009)	0.025*	(0.011)	0.025*
R Criminal Practice	-0.110**	(0.039)	-0.080*	(0.038)	-0.061	(0.034)	-0.178***	(0.043)	-0.178***
R is Attorney	-0.698***	(0.031)	-0.425***	(0.035)	-0.333***	(0.030)	-0.337***	(0.047)	-0.337***
Constant	4.915***	(0.111)	4.730***	(0.066)	4.711***	(0.067)	4.916***	(0.094)	4.916***
N (Judges)	4354 (71)		5887 (71)		5914 (71)		5906 (71)		5906 (71)
Adjusted R^2	0.02		0.06		0.06		0.05		0.05
Root MSE	0.896		0.775		0.663		0.920		0.920