
Are Judicial Performance Evaluations Fair to Women and Minorities? A Cautionary Tale from Clark County, Nevada

Rebecca D. Gill
Sylvia R. Lazos
Mallory M. Waters

Because voters rely on judicial performance evaluations when casting their ballots, policymakers should work diligently to compile valid, reliable, and unbiased information about our sitting judges. Although some claim that judicial performance evaluations are fair, the systematic research needed to establish such a proposition has not been done. By the use of attorney judicial performance survey data from Clark County, Nevada, this analysis shows that objective measures of judicial performance cannot explain away differences in scores based on race and sex. Minority judges and female judges score consistently and significantly lower than do their white and male counterparts, all other things being equal. These results are consistent with the hypothesis that judicial performance evaluation surveys may carry with them unexamined and unconscious gender/race biases. Future research must compare judicial performance evaluation structure, content, and execution across states in order to identify those evaluation mechanisms least susceptible to unconscious gender and race bias.

Ensuring that judges are accountable to the people is essential to maintaining the legitimacy of the judiciary. Accountability protects the independence of the judiciary (O'Connor 2009). This helps shield the judicial branch from the low public confidence ratings that are common in the executive and legislative branches. For these reasons, various judiciary-oriented organizations have been championing judicial performance assessments as an important tool in maintaining the judiciary's accountability to the people (Laube et al. 2007). These state judicial performance evaluation (JPE) programs are promoted for the pursuit of two goals: to provide information to the judges for the purposes of self-improvement and to provide information to the public about the performance of their judges.

State JPE programs should be "designed and administered in a way that does not inadvertently harm the principles they are

For helpful comments on previous versions of this article, the authors wish to thank Chris W. Bonneau, Eric Waltenburg, Udi Sommer, and Malia Reddick. Please address correspondence to Rebecca Gill, University of Nevada Las Vegas, Department of Political Science, 4505 S. Maryland Pkwy., Box 455029, Las Vegas, NV 89148-5029; e-mail rebecca.gill@unlv.edu.

Law & Society Review, Volume 45, Number 3 (2011)
© 2011 Law and Society Association. All rights reserved.

intended to promote” (Esterling 1998: 207). More specifically, JPEs should evaluate fairly the performance of all judges, regardless of their gender, race, or other personal demographic attributes. This is especially important because voters are influenced by performance evaluations (Esterling 1998). In judicial elections, one of the biggest causes of low voter participation is a lack of relevant information about the candidates (Bonneau & Hall 2009). Voters in most judicial elections get relatively little information about the candidates prior to the elections (Baum 2001).

In vigorously contested elections, citizens will learn about candidates through campaign advertisements. Voters in partisan elections see a voting cue in the form of party identification on their ballots. In retention elections, however, there is very little information available to voters besides the results of the JPEs (Esterling 1998). Exit polls of voters in four states using retention elections found that, “[a]mong those respondents who were familiar with the evaluation reports, most report that the evaluation information either determined or assisted their vote choice in retention elections” (Esterling & Sampson 1998:3). If the evaluation reports provided to voters are systematically biased based on gender or race, this will lead to less diversity on the bench as voters, in reaction to these evaluations, fail to retain women and minority judges.

The appraisal of judicial performance, like other types of performance evaluations, “is fraught with many potential pitfalls and problems” (Kearney 1999:483). Many state programs have struggled to create fair and workable programs (Pelander 1998). Nonetheless, evidence about the validity, reliability, and fairness of judicial evaluation programs is scarce (Kearney 1999). While many states have put in place judicial performance evaluation systems, none has conducted a systematic review to answer the basic question of whether these systems are valid, reliable, and fair.

This is a critically important issue at a time when the judiciary is still struggling to increase gender and racial representativeness. Women judges make up only a quarter of the federal bench and 32.4 percent of state supreme court judgeships, far below their representation in the legal profession (Commission on Women in the Profession 2009). As Supreme Court nominee Elena Kagan noted in an address about women in the legal field, “there’s a problem here, and we need to figure out why it exists” (Kagan 2006). Part of this problem may be related to role stereotyping. Early research on this topic showed that “[t]he image of the judicial office thus seems to affect negatively a female candidate’s chances” (Hedlund et al. 1979: 520). Interest groups continue to push hard for the transition to and reliance on these JPEs in merit-based selection systems. While some groups advocate “for strong judicial

performance evaluation (JPE) programs in every state” (Institute for the Advancement of the American Legal System 2010), they do so without the benefit of research to confirm that these programs will not contribute to systematic bias in the selection of our state judges.

Do judicial performance evaluations manifest unconscious bias through the stereotypes of the judicial role and of candidates for judgeships? This research is aimed at remedying this information gap. After a brief discussion of judicial performance evaluations in the American states, we present a theory as to why existing systems may be expected to yield biased results based on gender and race. Next, we present an analysis of judicial performance evaluations administered in Clark County, Nevada, from 1998 to 2008. Our results show that there is reason to question whether attorney polls reliably avoid problems of unconscious gender and race bias.

Assessing Judicial Performance with Surveys

Judicial performance evaluation is nothing new; polls of attorneys seeking opinions on judicial performance date as far back as the 1870s (Feeny 1987). In large part, the more recent movement toward state-mandated judicial performance reviews began in tandem with the shift toward merit plan judicial selection. These state-run evaluations arose as a reaction to the dearth of candidate information available to voters in retention elections (Pelander 1998).

In 1985, in anticipation of the growing need for accountability and independence of the judiciary, and with the first judicial performance evaluation program in Alaska nearly a decade old, the American Bar Association’s (ABA) Special Committee on the Evaluation of Judicial Performance published guidelines intended to direct states toward the “best practices” in the implementation of JPEs (American Bar Association 1985). Although the primary goal was to increase the ability of judges to improve themselves through constructive criticism and feedback, a secondary goal of the ABA report was to provide voters with valuable information on the quality of judicial candidates who were facing retention elections. Essentially, the ABA attempted to provide a roadmap for states seeking to create JPE programs by highlighting key criteria on which judges could adequately and appropriately be evaluated.

Since the issuance of the ABA’s guidelines, 23 states have created or are currently in the process of developing JPE programs for the purposes of voter information and awareness, judicial

self-improvement, and judicial education (Rottman 2004). Among other things, the ABA report laid out five main categories of criteria on which judges should be evaluated: legal ability, integrity/impartiality, communication skills, professionalism/ temperament, and administrative capacity. The first column of Table 1 presents the subcategories from the ABA report.

Among the states that have adopted some form of JPE program, nearly all utilize a survey as the key instrument to measure judicial quality (Pelander 1998). States go about the survey process in different ways. Surveys can be state-sponsored, bar association-run, and even privately sponsored. The newspaper-sponsored poll of the state bar that is the subject of analysis here, the *Las Vegas Review Journal's* (*LVRJ*) "Judging the Judges" survey,¹ is an example of the latter type. All of the state JPE surveys exhibit variations in terms of the groups to which they are distributed, the number of questions and the survey's contribution to the overall judicial evaluation process. However, all hew closely to the ABA guidelines of 1985 in terms of both the categories of information surveyed and the survey questions themselves. All survey-based evaluation systems currently in use canvass attorneys. State-sponsored programs tend to be more ambitious, and many also survey jurors, litigants, and/or other constituent groups. The officials compiling the data for the final evaluation, however, report that the lawyer polls provide the most helpful information (Esterling & Sampson 1998).

The widespread use of surveys to assess judicial performance stands in stark contrast to the limited use of objective information. Even though scholars (Kearney 1999) and judges (Institute for the Advancement of the American Legal System 2008)² agree that survey data should not stand alone as a determinant of judicial quality (Brody 2000: 341), few states formally integrate objective data alongside the survey data. Objective measures such as discipline records, caseload evaluations, reversal rates, recusal rates, and completion of continuing judicial education requirements provide a good picture of judicial performance (American Bar Association 2005; Andersen 2000). Unfortunately, states have found it difficult to collect and utilize these alternative measures of judicial performance (Thomas et al. 2009). This leaves most states relying very heavily or entirely on survey data alone.

¹ Information about the "Judging the Judges" survey can be found on the *Las Vegas Review-Journal* Web site: http://www.lvrj.com/hottopics/in_depth/judges/.

² According to the Institute for the Advancement of the American Legal System's 2008 survey of sitting Colorado judges, close to three-quarters believed that objective data, such as case management data, should be part of Colorado's JPE program.

Table 1. Comparing Questions between the ABA Guidelines, the *LVRJ* Survey, and Other State Surveys

ABA Guidelines	"Judging the Judges" Survey	Use by Other States
Legal Ability (5-1)		
1-1: Legal reasoning ability	The judge's rulings, whether regarding civil issues, criminal sentencing, or contempt are appropriate.	AZ, CO, HI, KS, NM, NC, NV, ^a WA
1-2: Knowledge of substantive law	The judge demonstrates familiarity with the case record and documents, and fairly weighs all evidence and arguments before rendering a decision.	AZ, CO, HI, KS, MO, NM, NC, NV ^a
1-3: Knowledge of rules of procedure and evidence	The judge properly applies the law, rules of procedure, and rules of evidence.	AK, AZ, CO, KS, MO, NM, NC, NV, ^a WA, NC, NV ^b
1-4: Keeping current on developments in law, procedure, and evidence	N/A	
Integrity and Impartiality (5-2)		
2-1: Avoidance of impropriety and the appearance of impropriety	The judge's professional conduct is free from impropriety and the appearance of impropriety.	AK, AZ, HI, KS, MO, NM, NC, NV, ^a WA
2-2: Treating all people with dignity and respect	N/A	AK, AZ, CO, HI, MO, NC, NV, ^a WA
2-3: Absence of favor or disfavor toward anyone, including but not limited to favor or disfavor based upon race, sex, religion, national origin, disability, age, sexual orientation, or socioeconomic status	The judge's conduct is free from bias on the basis of race, ethnic origin, gender, and religion (posited in several individual questions).	AK, AZ, CO, HI, KS, MO, NM, NC, NV, ^a WA
2-4: Acting fairly by giving people individual consideration	N/A	CO, KS, MO, NV ^b
2-5: Consideration of both sides of an argument before rendering a decision	The judge demonstrates familiarity with the case record and documents, and fairly weighs all evidence and arguments before rendering a decision.	AK, CO, KS, MO, NV ^b
2-6: Basing decisions on the law and the facts without regard to the identity of the parties or counsel, and with an open mind in considering all issues	The judge's conduct is free from bias on the basis of parties or attorneys involved in the action.	AK, CO, HI, KS, NC, NV ^b
2-7: Ability to make difficult or unpopular decisions	N/A	AK, CO, KS, MO
Communication Skills (5-3)		
3-1: Clear and logical oral communication while in court	The judge clearly explains the basis for his or her decisions.	AZ, CO, HI, KS, MO, NM, NC, NV, ^a WA
3-2: Clear and logical written decisions	The judge clearly explains the basis for his or her decisions.	AZ, CO, HI, KS, MO, NM, NC, NV, ^a WA

Table 1. Continued

ABA Guidelines	"Judging the Judges" Survey	Use by Other States
Professionalism & Temperament (5-4)		
4-1: Acting in a dignified manner	N/A	AZ, CO, HI, KS, MO, NC, NV, ^a WA
4-2: Treating people with courtesy	The judge is courteous.	AK, AZ, CO, KS, MO, NC, NV, ^a WA
4-3: Acting with patience and self control	N/A	AZ, KS, NC, WA
4-4: Dealing with pro se litigants and litigation fairly and effectively.	N/A	CO, NC
4-5: Participating and providing leadership to an appropriate degree in professional development activities and in jurisdiction-wide and statewide court improvement and judicial education activities	N/A	NV*
4-6: Promoting public understanding of and confidence in the courts	N/A	AZ, NV ^a
Administrative Capacity (5-5)		
5-1: Punctuality and preparation for court	The judge is punctual in convening court, keeps business moving, and does an amount of work fair to taxpayers and other judges.	AK, CO, HI, KS, MO, NM, NC, NV, ^a WA
5-2: Maintaining control over the courtroom	N/A	AK, CO, KS, MO, NM, NC, NV, ^a WA
5-3: Appropriate enforcement of court rules, orders, and deadlines	N/A	CO, KS, NM, NC, NV, ^a WA
5-4: Making decisions and rulings in a prompt, timely manner	The judge issues orders, judgments, decrees, or opinions without unnecessary delay.	AK, AZ, CO, KS, NM, NC, NV, ^a WA
5-5: Managing his or her calendar efficiently	N/A	CO, KS, NM, NC, NV, ^a WA
5-6: Using settlement conferences and alternative dispute resolution mechanisms as appropriate	N/A	CO, KS, NV ^a
5-7: Demonstrating appropriate innovation in using technology to improve the administration of justice	N/A	None
5-8: Fostering a productive work environment with other judges and court staff	N/A	None
5-9: Utilizing recruitment, hiring, and promotion policies and practices to ensure that a pool of qualified applications for court employment is broad and diverse	N/A	None
5-10: Acting to ensure that disabilities and linguistic and cultural differences do not limit access to the justice system	N/A	None

^aNevada's proposed state-sponsored JPE program would have been implemented if SJR-2 had passed in November, 2010 (see Thomas et al. 2009).

The Potential for Race and Gender Bias in the Survey Component of JPEs

The limited use of objective data in evaluating judges is troubling because objective metrics could help counter two types of systemic errors that arise from the use of survey data. First, JPE surveys often contain methodological errors that can be traced back to lack of professional guidance and failure to incorporate statistical controls (Brody 2000: 341).³ Second, judicial performance surveys, like teacher evaluations (Sprague & Massoni 2005) and other job performance assessments (Martell 1991), could be subject to unconscious gender and race bias. In this analysis, we focus on the possibility of the latter type of systemic error.

Judges have long expressed concern about the fairness of evaluation systems (Griffin 1994; Institute for the Advancement of the American Legal System 2008). One main concern is the possibility that unconscious gender and race bias may color the recommendations of the judicial evaluation committees. Of the 22 states that have state-sponsored JPE systems in place, not one has undertaken the work of systematically reviewing judicial performance surveys for evidence of such unconscious gender and race bias. As other researchers have noted, “there is a self-evident need to acquire more factual data about the operation of gender in the judicial performance evaluation process” (Durham 2000: 16). The same must be said for racial bias.

Conscious or explicit bias is the kind of prejudice that people openly embrace, but unconscious or implicit biases are automatic, unintentional, and often unexamined. This latter type of bias can cause people who believe they are neutral, and who espouse principles of race and gender equity, to nonetheless make erroneous judgments influenced by race, gender, or other stereotypes (Lane et al. 2007). Even psychologists (Steinpreis et al. 1999) and trial court judges (Rachlinski et al. 2008) have been found to exhibit this type of unconscious bias.

Within the discipline of psychology there has been extensive research establishing that stereotyping, as part of the normal unconscious cognitive processes of categorization, leads to inaccurate and unfair judgments of women and minorities (American Psychological Association 1991). A stereotype can be defined as a “set of attributes ascribed to a group and imputed to its individual members simply because they belong to that group” (Heilman

³ Judges themselves are concerned with the possibility that JPEs are based on unreliable survey data. According to the 2008 IAALS survey of sitting Colorado judges, a majority of judges perceived that low response rates were a “major problem” in the survey component of JPEs (Institute for the Advancement of the American Legal System 2008).

1983). Stereotypes act as cognitive shortcuts that form the bases for cognitive categorizations, leading us to generalize about individuals and circumstances. For our purposes, racial and gender stereotypes are negative generalizations based on culture and myth. Social norms rooted in the past justified racial and gender hierarchy based on “truths” about gender and race. These “truths” disadvantaged the minority status group and legitimized the dominant group’s privileged status. Although we have moved to a more equitable society, these stereotypes linger and continue to affect how we perceive those in the minority group.

Recent research shows just how strongly stereotypes influence the actions of actors who believe themselves to be without prejudice (Bertrand & Mullainathan 2004; Correll et al. 2007). Within the body of anti-discrimination law, the powerful effect of these unconscious biases in the evaluation process has long been recognized. For example, in *Price Waterhouse v. Hopkins* (1989), the U.S. Supreme Court found that the evaluation of a female candidate for partner in an accounting firm was likely driven by unconscious bias. Researchers have found that women’s performance ratings in jobs that are sex-typed as male (e.g., police officers) suffer when they are evaluated in comparison with men’s ratings, and vice versa (Heilman et al. 2004; Nieva & Gutek 1980).

Women judges perform work stereotypically considered to be “a man’s job.” Women in male-typed jobs are not conforming to gender role expectations (Valian 1998). This theory predicts that women judges will be at a disadvantage, failing to be recognized and rewarded when they are doing a good job. Certainly, some individual female judges will be able to defy stereotypes, but, by and large, most women are not able to pull off such a performance. Many women judges have to work harder and do more things right to get a score that would still be a low score for male judges.

Similar biases related to race-based stereotypes, especially in the context of the law-and-order world of judging, may drive inequities in retention scores based on race. Trial judges may exhibit unconscious race bias in their own work (Rachlinski et al. 2008), and stereotypes surrounding the racial characteristics of criminal defendants may exacerbate the racial disparities in evaluating minority judges. Social science research shows that, while the underlying racial stereotypes are different from those underlying gender bias, “the consequences of skewed racial distributions for the social psychology of stereotyping and outgroup bias are similar to those resulting from gender imbalance, as are the resulting barriers to career advancement” (Bielby 2000: 123). When a candidate’s true qualifications are somewhat ambiguous, minority applicants are consistently given less support than their white counterparts (Dovido & Gaertner 2000). Minority evaluatees are

Table 2. Descriptive Statistics and Difference of Means Tests by Gender and Race

	Mean	SD	Female Mean Difference [†]	F	Minority Mean Difference [‡]	F
Retention score	73.84	16.66	-10.59	9.48**	-12.57	2.75
Reversal rate*	0.08	0.09	-0.01	0.43	-0.10	0.33
First appointed	0.42	0.50	-0.02	0.03	0.62	7.88**
Law school tier	1.98	1.17	0.66	7.03**	-0.19	0.12
Years' experience	7.83	7.08	-2.27	2.24	-1.95	0.35
Discipline result	0.54	1.71	-0.40	1.21	0.91	1.35
Reported scandal	0.21	0.41	-0.04	1.82	0.53	0.27

N = 95 judges (33 female and 5 minority), weighted by observations for each judge in the data (total = 364)

*The independent variable in the model has been weighted by the judges' bench means. There is no significant difference between these weighted scores based on gender.

[†]This number is calculated by subtracting the mean for male judges from the mean for female judges. Where the sign is negative, the mean value for female judges is lower.

[‡]This number is calculated by subtracting the mean for white judges from the mean for minority judges, as above. Note that there are only 5 minority judges in the sample.

judged more fairly when they compose at least 30 percent of the pool (Sackett et al. 1991), but minority judges in the "Judging the Judges" survey pool represent just half of 1 percent (see Table 2).

Many states have recognized the potential for unconscious bias—especially gender bias—in the judicial system. To date, 45 states have implemented some type of task force charged with evaluating the level of gender bias in state court systems (Schafran & Wikler 2001). These task forces are generally asked to determine the severity and type of bias. Many of them have found evidence of substantial gender bias in courtrooms, offices of court administration, and in the legal profession more generally (Kearney & Sellers 1996). Unfortunately, gender bias in the selection process is "still a problem" (Schafran & Wikler 2001: 34).

The most highly publicized debate about possible gender and race bias in judicial performance evaluations occurred in Missouri in 2007, when revisions were made to the attorney poll that had been administered by the Missouri Bar since 1948 (Missouri Bar Judicial Evaluation Survey Committee 2007). The revisions were motivated by a desire to improve the survey, and they reflected the Bar's hope that these improved judicial evaluations would have greater influence on voter behavior.⁴ In 2004, the Missouri Bar revamped the survey, and the process resulted in a longer, more comprehensive questionnaire. The new survey instrument more

⁴ In 1990, according to the Missouri Bar Report, "it became apparent that the election results did not mirror the legal community's perception of the judges' performance, raising a concern that the public did not have sufficient knowledge and understanding of the judges' qualifications" (Missouri Bar Judicial Evaluation Survey Committee 2007).

closely conformed to the 1985 ABA guidelines.⁵ Surprisingly, this more detailed survey resulted in greater scoring gaps by gender and race. By moving from a simple “yes/no” retention question to a more sophisticated survey structure, Missouri increased the evaluation disparities by gender and race. The new survey resulted in negative recommendations for two associate circuit court judges who were black females, each of whom had previously received “do retain” ratings (Lauck 2006).

At the request of the Mound City Bar Association, Dr. Gary Burger undertook an analysis of Missouri’s judicial performance evaluations for a 10-year period (Burger 2007). The Burger analysis found that there was a significant difference in the Missouri attorney poll results according to gender and race. Female judges, as a group, were ranked much lower than male judges, and black judges ranked lower than white judges (Burger 2007).⁶ Of the four groups that could be formed by classifying judges by both gender and race, female African-American judges were rated significantly lower than white female judges, and African-American male judges were rated statistically significantly lower than Caucasian male judges (Burger 2007).

Burger’s study does not ascribe a reason for these significant differences in ratings. In fact, he specifically states that “[t]here are a variety of possible explanations for the differences documented in this study . . . [such as] bias on the part of the raters, real differences in performance, problems with the rating scales and process, or some combination of these factors” (Burger 2007: 6). It is clear that a controlled analysis is needed in order to provide evidence as to which of these factors is driving score disparities in JPE surveys.

Data and Methodology

The purpose of this study is to remedy this research gap. We supplement attorney survey results with alternative measures of judicial quality in order to test the hypothesis that unconscious gender and race bias are influencing bar poll results. The dependent variable in our analysis comes from data from the *Las Vegas Review-Journal*’s “Judging the Judges” biennial survey (1998–2008)

⁵ The ratings categories increased from 5 to 16. Attorneys were asked to detail whether a judge gave reasons for ruling, engaged in *ex parte* communications, made correct decisions based on law and facts, had settlement skills, and demonstrated impartiality, decisiveness, courteousness, and fairness (Burger 2007).

⁶ In the Burger study, male judges were rated significantly higher (85.84 percent) than were female judges (76.69 percent). Caucasians were rated significantly higher (85.44 percent) than were African Americans (75.19 percent). This ratings gap was present in all three courts studied.

of local attorneys rating Clark County judges and Nevada's Supreme Court justices. Until 1996, this survey was conducted jointly with the local bar association. Since 1998, the "Judging the Judges" survey has been conducted solely by the *Las Vegas Review-Journal*, a private newspaper, with the help of an external research firm. The *LVRJ* makes a redacted version of the survey results public on their Web site.⁷

The surveys cover all of the judgeships that appear on the Clark County ballot. This includes Clark County trial judges and the justices sitting on the Nevada Supreme Court.⁸ In all, there are 95 judges in this dataset, 33 of whom are female and 5 of whom are of minority racial status. Because many judges are rated in successive biennial surveys, there are a total of 364 observations. The eligible pool of respondents consists of all attorneys who hold a bar card in that jurisdiction (Whitely 2006). The number of respondents for each survey reaches as many as 800 lawyers. Each attorney is asked to affirm that he or she has had direct experience dealing with the sitting judge that he or she is rating (Hopkins 2008), although no validation process is in place to ensure that attorneys limit their responses to only judges before whom they have appeared during the evaluation period.

Our data include the universe of judicial evaluations from these biennial surveys from 1998 to 2008. The survey includes the questions outlined in the second column of Table 1.⁹ These questions address most of the categories recommended in the ABA's guidelines (American Bar Association 1985). The last question in the survey reads, "Taking everything into account, would you recommend retaining this judge on the bench?" The percentage of respondents answering "yes" to this question represents the retention score. This is the dependent variable in our analysis.¹⁰ As Table 2 shows, the average retention score in these data is 73.84 of a possible 100.

⁷ The redacted form of the survey results from 2000–2008 can be found at the *LVRJ*'s Web site: http://www.lvrj.com/hottopics/in_depth/judges. This version reports aggregate results only. Attorney comments are also redacted, although the newspaper has selectively included these attorney comments in their reporting.

⁸ Nevada currently has no intermediate appellate court.

⁹ Between the 2000 and 2002 survey, the questionnaire was shortened from 15 questions to 13 questions. The new, shorter version did not omit questions, but instead collapsed some of the questions.

¹⁰ In some of the survey reports, an "adequacy" score was also provided. The questions on the survey, with the exception of the retention question, asked attorneys to rate judges on a three-point scale: not adequate, adequate, or more than adequate. The adequacy score was calculated by taking the percentage of "adequate" or "more than adequate" ratings the judges received. We ran our model with this adequacy score as the dependent variable, and the results were quite similar to what is presented here.

Table 1 presents a comparison between the ABA guidelines, the “Judging the Judges” bar poll and the attorney surveys used in other states. There is remarkable uniformity regardless of which organization is sponsoring the survey. States generally hew to the guidelines issued by the ABA. The greatest variability among these is in the length of the surveys; for example, Colorado’s survey of district attorneys is composed of 36 questions.¹¹ Although many states use surveys that are significantly longer than the “Judging the Judges” bar poll, there is no evidence that longer surveys produce better evaluative results (Burger 2007). This comparison shows that the results of this analysis of the *LVRJ*’s “Judging the Judges” bar poll may be generalizable to state-sponsored programs elsewhere.

The independent variables of interest are gender and minority status. A dummy variable is included to indicate whether the judge is female, and another indicates whether the judge is of a minority race. Similar to the findings of the Burger study (2007), the retention score exhibits a consistent, statistically significant difference among judges based on both gender and race. Table 2 shows the frequency of votes in favor of retention broken down by gender and race. The first line of Table 2 shows that the mean for male judges is significantly higher—by more than 10 points—than for female judges. The mean difference between minority and nonminority judges is even bigger, at more than 12 points. This difference is not statistically significant, perhaps because there are only 5 minority judges in our sample. The distribution of scores for male judges is also skewed, with more high scores than would be predicted by the normal distribution (see Figure 1).

It is important, of course, to rule out the possibility that this difference in ratings is attributable to female and minority judges being, on average, less qualified and less able on the bench. For this reason, a number of alternative indicators of judicial performance are included in the model. If there is a qualitative difference in the abilities of women judges versus male judges and minority versus nonminority judges, then the gap in scores represents a true difference in judicial performance. In that case, the observed differences would be attributed not to bias, but to true differences in these alternative measures of judicial quality. In order to assess the relative quality of the judges based on gender and race, we included proxy measures for the quality of the judges in the study; specifically, we include weighted reversal rates for each judge, whether a judge was initially appointed to the bench, quality of legal education, experience on the bench, and ethical record. The

¹¹ Colorado’s survey does not present new questions or categories in terms of concepts. Essentially, Colorado is not asking new questions, but simply fracturing existing ones (Colorado Office of Judicial Performance Evaluation 2008).

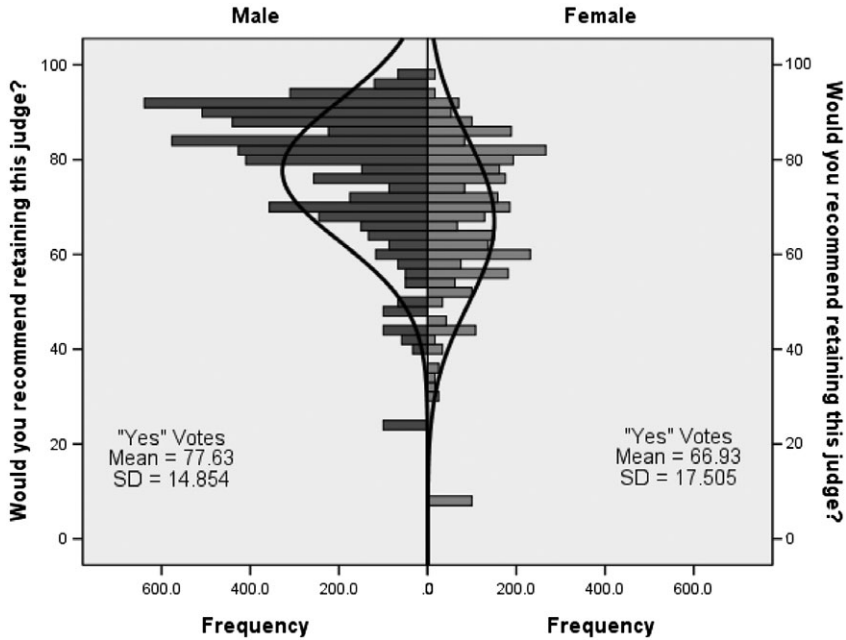


Figure 1. "Yes" votes for retention in LVRJ Bar Poll Survey by gender, 1998–2008.

descriptive statistics for these variables are included in Table 2, along with the results of difference of means tests by judge gender and minority status.

Reversal Rates

One widely used source of objective information about legal ability is the reversal rate (Brody 2000; Posner 2000). When a judge has a low reversal rate, this is an indication that the judge is interpreting the law correctly in the eyes of the appellate court (Feeny 1987). For this study, we gathered reversal rate figures using a Lexis–Nexis search for each judge spanning the years 1998–2009.¹² Table 2 shows that there is no significant difference in reversal rate by judge, gender, or race. Because our sample includes judges from across the spectrum of jurisdictions, the measure of reversal rate included in the model is calculated as

¹² This yields a single reversal rate for each judge in the database, and this rate spans his or her time as a judge in Nevada. For example, a supreme court judge's reversal rate includes reversals from decisions made in his or her previous judicial posts.

the distance between an individual judge's reversal rate and the mean for his or her court. The mean value of this variable is zero, and this represents the mean reversal rate for the level of court. In our dataset, the reversal rate distance ranges from -0.35 (which is a lower or "better" reversal rate than the court mean) to 0.13 .

We must entertain the possibility that the reversal rate itself is partially driven by race and gender bias. Attorney prejudice or bias as to the reversibility of the judge's decision may influence the attorney's decision as to whether there is a better than average probability that the judge will be reversed on appeal. If attorneys have a low opinion of a female or minority judges' legal abilities, reflecting the attorneys' own biases, then female and minority judges' rulings may be more often subject to appeal. The higher frequency of requests for appellate review may result in higher reversal rates, regardless of whether the individual judge is competent. As Walker and Barrow observe, "if practicing attorneys perceive [female and black judges] to be vulnerable to reversal on appeal, then a pattern of relatively high appeal rates for nontraditional judges should occur" (1985: 612).

To test this possibility, difference of means tests were conducted on the both the reversal rate distance and the appeal rate distance. The appeal rate distance is calculated the same way as the reversal rate distance, but using the number of times a judge was appealed as compared with the mean of the relevant court. Because this measure necessarily uses a raw number as opposed to a rate, it is weighted by the number of evaluation years we have for each judge. The difference of means tests reveal no significant bias effects in either of these variables.¹³ This is in keeping with findings of previous research on diversity on the federal bench (Walker & Barrow 1985).

Initially Appointed to the Bench

Another proxy measure for judicial quality is whether or not the judge was first appointed to the bench. Nevada's constitution provides that the justices and judges who serve in the courts of Nevada shall be elected (Nev. Const., Art. VI sec. 2, 3A). Unlike most other states with competitive judicial elections, the Nevada constitution stipulates that judicial vacancies be filled by the governor from a roster of recommended nominees selected by a com-

¹³ The results of the difference of means tests for female judges (on reversal rate, $F = 0.487$; on appeal rate, $F = .515$) and for minority judges (on reversal rate, $F = 1.003$; on appeal rate, $F = 0.089$) yielded no statistically significant differences in either reversal rates or appeal rates.

mittee of citizens, which is largely made up of lawyers.¹⁴ So, in the case of vacancies only, Nevada employs a merit selection plan of sorts (Flango & Ducat 1979). Empirical studies have found that having been appointed at first instance confers benefits on judges; most notably, they gain incumbency status without first having to stand for election (Glick 1978). This suggests that, from the voter's perspective, judges who are appointed in the first instance have desirable qualities that make them worthy of being judges.

The evidence is mixed as to whether judges selected through a merit plan are of higher quality than elected judges. In fact, there is empirical work that concludes that appointed judges are *not* qualitatively superior to elected judges (Bonneau & Hall 2009; Choi et al. 2007; Glick & Emmert 1986). Regardless of the empirical evidence, attorney respondents may believe that appointed judges are more highly qualified. First, attorneys tend to support merit plans as the best method of selecting judges (O'Connor 2009). This may be because the state bar plays such a critical role in the selection process. Lawyers usually dominate the merit selection committees that make recommendations to governors, and state bars have great influence on who gets to sit on these committees. Recommended nominees already have close connections with the state bar and would be highly regarded by members of the bar. In other words, because the bar has a hand in choosing merit-selected judges, attorneys may be inclined to see these judges as more highly qualified than those selected by popular election.

Second, there is an argument that in nonpartisan elective systems like Nevada's, judges initially appointed to the bench may indeed be of higher quality than elected judges. Elected judges in nonpartisan systems face no vetting process prior to taking their place on the ballot. In Nevada, the only constitutional and statutory requirements to run for some judgeships are residency, minimum age, and collecting sufficient signatures on a petition.¹⁵ The absence of a vetting process (which, in partisan elections, is often provided through the political party system) may indeed lead to lower-quality candidates.

¹⁴ Nev. Const. Art. VI, Sec. 20(1). In most states, vacancies are filled by gubernatorial appointment as opposed to a merit-style plan.

¹⁵ Nev. Rev. Stat Sec. 5.020. Additional statutory requirements for justices of the peace include a high-school diploma and, in some jurisdictions, membership in the state bar association (Nev. Rev. Stat Sec. 4.010). Additional statutory requirements for District Court judges include a minimum of 10 years as a member of the bar in an American state or the District of Columbia, 2 years of which must have been in Nevada (Nev. Rev. Stat Sec. 3.060). Additional statutory requirements for the justices of the Nevada Supreme Court include a minimum of 15 years as a member of the bar in an American state or the District of Columbia, 2 years of which must have been in Nevada (Nev. Rev. Stat Sec. 2.020).

In this study, we include a dummy variable to indicate whether the judge was appointed to the bench on first instance. Table 2 shows that there is no significant difference between men and women in terms of appointment rates. The mean indicates that 42 percent of the judges in this sample were appointed on first instance. In our sample, all five of the judges with minority race status were first appointed to the bench.

Legal Education

The nature of a judge's legal education is another proxy for judicial quality, "though like most proxies a rough one" (Posner 2005). The prestige of a judge's legal education carries with it certain implications for the judge's perceived level of legal ability (Glick 1978).¹⁶ Research on the American Bar Association's ratings of federal judicial nominees indicates that law school prestige is an important covariate of the ABA's quality scores (Slotnick 1983). For this analysis, we use a variable representing the current tier ranking of the judge's law school education as per the 2010 rankings from *U.S. News & World Reports*.¹⁷ Alumni information was gathered from various online sources, including official judge biographical blurbs, résumés, and newspaper articles.

Of the 95 judges in this dataset, 36 came from the very best schools and 34 came from second-tier schools. On the lower-prestige end, 6 judges came from tier-three schools, 16 from tier-four schools, and three did not graduate from law school at all. Table 2 shows that the average level of law school prestige is a second-tier school. The prestige of school attended by the female judges is, on average, about three-quarters of a tier lower. There is no significant difference between minority and non-minority judges, but it is interesting to note that minority judges have a mean that is about one-fifth of a tier higher than their non-minority peers.

Judicial Experience

Previous judicial experience is a time-tested measure of judicial quality, particularly with respect to the federal bench (Epstein et al. 2003). In the realm of judicial quality, it stands to reason

¹⁶ Glick (1978) also suggests that the in-state/out-of-state law school distinction is important. In Nevada, this distinction is less helpful. Nevada's first law school, the William S. Boyd School of Law at the University of Nevada Las Vegas, did not graduate its first class until 2001.

¹⁷ These rankings can be found online at <http://grad-schools.usnews.rankingsandreviews.com>.

that a judge who has served longer has had more opportunity to learn and master the trade of judging (Haire 2001). Research on federal court judges suggests that recently appointed judges experience acclimation effects as they familiarize themselves with the job of judging (Brenner & Hagle 1996; Hettinger et al. 2003). Of course, it is possible that the longer one stays on the bench, the more likely it is that the respondent attorneys will see them as “old-timers,” or otherwise past their prime. This is certainly part of the argument for imposing a mandatory retirement age on judges.

As with the information about law schools, judicial experience information was gathered from publicly available sources. Judges in our sample ranged from 0 to 30 years on the bench. The average tenure of judges in our sample is nearly 8 years. Female and minority judges had slightly shorter tenures, on average, but this difference is not significantly significant (see Table 2).

Ethical Record

Disciplinary complaints are an important signal about a judge’s performance (Posner 2005), and may be the single most important objective measure of judicial integrity available to researchers. To be sure, some lawyers may be unwilling to file formal complaints for fear of future retribution (Jackson 2007); however, filing a disciplinary complaint is also relatively easy. We have elected to code a disciplinary record variable to reflect both whether a complaint was filed and the outcome of the disciplinary complaint, rather than just the filing of the complaint. These outcomes were ranked on an ordinal scale of 0 to 8.¹⁸

We have also included a separate variable that reflects reported scandals and accusations of impropriety. Not all scandals and press reports alleging that a judge has behaved improperly rise to the level of disciplinary complaints. However, some respondents may follow the philosophy of “where there is smoke, there is fire.” Even the mere suggestion of impropriety may drive attorney opinion against a judge who is the subject of public reports of misconduct. Alternatively, respondents may be swayed by a judge’s connection to public scandal as, at the very least, unbecoming of a member of the judiciary. Attorneys without recent, first-hand experience with

¹⁸ The scale for this ordinal variable is as follows: 0 = no complaint; 1 = complaint was filed but dismissed; 2 = required course; 3 = required course and public apology; 4 = public reprimand; 5 = public reprimand and fine; 6 = censure, required course, and fine; 7 = removal from bench; 8 = removal from bench and permanently barred from holding public office in Nevada.

the judge may be particularly susceptible to being swayed by this sort of information.¹⁹

For this reason, we have included a dummy variable to flag those judges who have been publicly connected—rightly or wrongly—to scandal. This information comes from publicly available news sources, mostly from the pages of the *Review-Journal* itself. In our study, a full 21 percent of the judges are involved in a reported scandal (see Table 2). This high number may reflect Clark County's reputation for being one of America's most preeminent "judicial hell-holes" (American Tort Reform Foundation 2008). Alternatively, it may reflect the fact that the media in Las Vegas is particularly vigilant in reporting the actual and alleged missteps of elected judges. As Table 2 demonstrates, there is no significant difference based on gender or race on this measure.

Methodology

An important consideration in an analysis of this sort is the relative weight given to the various judge ratings in the *LVRJ* poll. While the response rates of this poll often fall below the generally accepted level of 50 percent respondent solicitations (Brody 2000), the assessment of response rate is actually quite a bit more complicated than it seems. In the *LVRJ* methodology, like many of the other state surveys, the respondents are instructed to answer questions only if they have professional experience with the judge. In 2008, for example, researchers sent out 4,237 survey invitations, but only 799 attorneys responded (Downey 2008). This yields an incredibly low response rate of just 18.9 percent. Beyond this, none of the responding attorneys had experience with all of the 68 judges evaluated in the 2008 survey. For the 2008 results, an average of 202 attorneys evaluated each judge, with a range between 44 and 387. Given the limitations on the information about which attorneys actually had professional experience with each judge during the evaluation period (Thomas et al. 2009), it is impossible to calculate a "true" response rate of qualified attorneys. Even still, the results calculated with the input of fewer attorneys will be less reliable. As such, a weighted model of retention scores is used here.

Our dataset includes biennial data for Clark County judges and justices of the Nevada Supreme Court from 1998–2008. Of course, most of the judges in the dataset served on the bench for only some

¹⁹ The Nevada bar poll, like other attorney polls, asks respondents to evaluate only those judges that attorney-respondents have appeared before during the evaluation period. However, many judges in Nevada and elsewhere report that the number of respondents who rate them in attorney polls is *greater* than the number of attorneys that appeared in their courtroom.

Table 3. Pooled OLS Model * of *LVRJ* “Judging the Judges” Retention Scores

	Coefficient	Standard Error	z-score
Constant	89.24	(1.65)	53.03***
Reversal rate	-4.37	(4.45)	-0.98
First appointed	1.76	(0.60)	2.95**
Law school tier	-0.56	(0.39)	-1.43
Years' experience	-0.39	(0.15)	-2.66**
Discipline result	-1.94	(0.91)	-2.14*
Reported scandal	-5.65	(0.62)	-9.18***
Female judge	-11.27	(3.66)	-3.08**
Minority judge	-14.27	(1.59)	-8.98***

*Pooled ordinary least-squares regression with Driscoll & Kraay (1998) standard errors. $N(\text{obs}) = 364$; $N(\text{judges}) = 95$; $F(8,93) = 1980.93***$; root mean squared error = 316.73.

of this time. This means that we have an unbalanced panel dataset. To analyze these data, we estimate a pooled weighted least-squares (WLS) model with Driscoll and Kraay (1998) standard errors. This adaptation of the Beck and Katz (1995) approach allows for the regression to be adapted on the basis of a robust estimate of the error structure of the model (Hoechle 2007).²⁰ As we have several biennial estimates for most judges, we cannot assume that these observations are independent of one another. This model allows the standard errors to be calculated with this systematic dependence in mind.

Results

The results of our analysis are presented in Table 3. None of the objective measures of judicial quality and judicial performance that we utilize in our model mitigate the difference in scores based on race and sex. There remains a large, unexplained gap in the ratings of female and minority judges and their male and nonminority counterparts, all other measures of judicial quality being equal. This is demonstrated by the large, statistically significant coefficients on the race and gender variables in the model. These results are consistent with the hypothesis that judicial performance evaluation surveys may carry with them unexamined and unconscious gender and race biases (Durham 2000).

In the model, only two of the control variables fail to help explain variation in retention recommendation rates. The prestige of law school education is insignificant, as is the reversal rate distance from the court mean. Part of the reason may be that, as compared with

²⁰ The models in this paper were also estimated without the number of respondents as a weight, and again using the original Beck and Katz (1995) panel corrected standard errors method. The results of all of these analyses were similar to what is reported here.

measures of integrity, legal knowledge questions make up a small fraction of the questions on the *LVRJ* survey. Because the insignificance of these legal knowledge variables is interesting in its own right, it will be addressed separately in the discussion section below.

The rest of the variables are statistically significant in the model. The intercept for this model is about 89 points, which is the estimated starting place for favorable attorney retention ratings when all other variables are held constant. Judges who are appointed at first instance get a boost in scores of 1.76 points. Substantively, this is not of a particularly high magnitude, and it may reflect the legal profession's general preference for an appointment system of judicial selection (Standing Committee on Judicial Independence 2000). In 2010, 68 percent of responding Clark County lawyers favored a move to a merit plan for judicial selection (McMurdo 2010). Indeed, those appointed on first instance are able to achieve incumbency without first having to withstand the rigors of a contested election, and this may also help to protect their image among attorneys.

The number of years of experience on the bench is significant and negative. For each additional year of judicial tenure, the judge loses more than a third of a point. Lawyer respondents tended to disfavor retention at a higher rate for judges with more years accumulated on the bench.²¹ The integrity measures are also statistically significant. As a judge's disciplinary result moves one step up the scale of severity, the judge can expect to lose about two points. When the judge is connected with a public scandal, he or she can expect a loss of 5.65 points.

Of course, the independent variables of interest are both statistically significant and of very high magnitude. When all of these control variables are equal, we can expect female judges to score 11.27 points lower than their male colleagues. Minority judges score more than 14 points lower than nonminority judges. Even after controlling for the alternative measures of judicial quality, these judges' demographic variables still account for much of the variation in retention scores.

Discussion

The results of this analysis of the *LVRJ* "Judging the Judges" bar poll shows that the gender and race disparity in retention

²¹ Because the model showed no significant relationship between experience and retention scores, a nonlinear version of this relationship was also estimated, operating on the premise that judicial experience may have diminishing returns over time. In other words, we tested the hypothesis that some experience would increase scores, but judges who had been on the bench for decades would be penalized for being "too old" or "out of touch." We found no evidence to support this hypothesis.

scores is not driven by qualitative differences in judicial performance. If real performance differences were responsible for the gender- and race-based differences in scores, much of this disparity would fall away with the inclusion in our model of the objective measures of performance quality. This, however, is not the case. While many of the measures were significant, they failed to cancel out the gender and race disparities.

Interestingly, two of the more theoretically compelling measures of judicial quality were completely insignificant in our model. First, we find that the prestige of a judge's legal education is not related to the retention scores in the "Judging the Judges" surveys. This contradicts previous research about judicial quality (Glick 1978; Slotnick 1983). One explanation is that the key skills that judges need to be good judges are not necessarily learned in law school. It may be that the attorney respondents have a more sanguine assessment of the usefulness of traditional legal education to the art of judging. Law school education has been widely criticized for failing to address the tools needed by attorneys and judges (Sternlight 1996). While law schools do a good job in teaching legal theory, they do a relatively poor job training students in key practical skills that they would need in legal practice (American Bar Association 1992). So, perhaps the reason why legal practitioners find law school pedigree to be noninfluential in their ratings is that standard legal education does not teach the skills that are most relevant to being a good judge. Indeed, "[d]espite the fact that law school graduation is the path to judgdom, American law schools devote not one minute of education to preparing law graduates to enter the world of judging" (Cordell 2008:639). Even still, this does not explain why previous research on state and federal judges has found strong ties between the prestige of a judge's legal education and his or her subsequent quality ratings (Glick 1978; Slotnick 1983).

Reversal rates were also insignificant in our model of retention scores. Judges who make more "correct" decisions are not rewarded for it by the respondent attorneys in these surveys. In theory, reversal rates are among the most relevant and quantifiable objective measures we have. That they are not related to retention scores in the "Judging the Judges" survey shows that Clark County attorneys do not place much weight on a judge's reversal rates. This result may reflect the particularities of Nevada. It is one of only 10 states without an intermediate appellate court. Of these states,²² Nevada is by far the most populous. Because all appeals are heard by the Nevada Supreme Court, the chances of a case being heard

²² These states include Delaware, Maine, Montana, Nevada, New Hampshire, Rhode Island, South Dakota, Vermont, West Virginia, and Wyoming.

on appeal are relatively low. Nevada's 77 trial courts dispose of close to a half million nontraffic cases on a yearly basis (Titus et al. 2009). In 2009, the number of cases that were filed for appeal and were granted was 1,759, about 0.4 percent of all cases (Titus et al. 2009). A recent Department of Justice analysis of 48 large counties showed an appeal rate of nearly 15 percent in civil cases alone (Cohen 2006). Rates of criminal appeals tend to be higher still, with as many as one in five criminal cases resulting in an appeal (Baum 2001). Because cases in Nevada have a lower probability of being appealed in the first place, Nevada's lower court judges have a lower overall probability of being reversed than do trial court judges in other jurisdictions, all other things being equal.

Those alternative measures of judicial quality that are significant in the model also present some interesting stories. The two measures of ethical integrity—the presence of a scandal and the outcome of disciplinary complaints—have a high magnitude of impact on attorney ratings. These results, taken together, reflect the importance of perceived judicial integrity to attorney ratings of judges. There exists a long-standing ethic of judicial neutrality and integrity in the United States, and there is an expectation that judges will act in an unbiased, fair and impartial manner (Committee on Judicial Independence 2006; Brody 2008). For judges to be able to administer justice, they must be perceived as neutral parties who can render judgment without bias. The preamble of the Model Code of Judicial Conduct states that a judge “must respect and honor the judicial office as a public trust and strive to maintain and enhance confidence in the legal system” (American Bar Association 2007:1). In the recent case of *Caperton v. Massey* (2009), the U.S. Supreme Court found that the appearance of improper ethics rises to the level of a constitutional due process violation.

Additionally, reported improprieties are influential on attorney ratings above and beyond the actual adjudication of an ethical violation. This may reflect the fact that press information about alleged scandals is more current and plentiful. Because of the wide reach of media reports, these reported scandals are likely to be more salient to attorneys evaluating a judge's judicial ethics. Although there is a significant amount of overlap between disciplinary actions and media reports, successful formal complaints against a judge will underreport whether a given judge lacks integrity. Press reports will be censored only in terms of whether the newspaper deems the information newsworthy, and will often be reported in more memorable and sensational terms. By the time a judge is disciplined, which generally happens about a year from when the complaint was filed, the judge's improprieties are “old news” and the reputational market among attorneys would have already accounted for the misdeed.

The main story here, however, is the performance of the independent variables of interest: race and gender. These variables are both statistically significant and of very high magnitude. When all of the alternative measures of judicial quality are equal, we can expect female judges to score 11 points lower than their male colleagues, and minority judges score 14 points lower than nonminority judges. These differences, then, are not due to a qualitative difference in the abilities of women and minority judges, at least not in the alternative measures included here.

The hypothesis that the difference in means based on race and gender in this attorney poll is due to differences in talents and competence is not a convincing explanation for the systematically lower retention scores of female and minority judges. Even when these judges are on par with their peers in terms of experience, education, and integrity, their scores are still drastically lower than those of their white male counterparts. Of course, one could still make a case that our study has not entirely eliminated the possibility that the gap in ratings is due to true differences in talents and competency. We concede that there are alternative objective measures of judge quality not included in our model (e.g., Choi et al. 2009). Unfortunately, it is exceedingly difficult to obtain additional quality data reliably in state judiciaries. This issue is an appropriate subject for further inquiry and innovation.

Nonetheless, a significant finding of this study is how little effect our alternative measures of judicial quality have on attorney assessments of a judicial performance in the survey of Clark County attorneys. If it were true that the gap is a result of real qualitative performance differences, we would expect the objective measures in our model to explain much of the variance in retention scores. This is not what we find. The control variables, which represent this alternative hypothesis, do not overtake in magnitude the effects of gender and race in the model. This result illustrates the need to examine more closely the content and structure of existing JPE attorney surveys, as well as the central role of these surveys in determining committee retention recommendations in state-sponsored JPEs. While the differences between state surveys may mitigate some of the bias found in the "Judging the Judges" survey, empirical evidence must be amassed to support this conclusion.

If we reject the hypothesis that the difference in ratings is due to an inferior pool of talent among women and minority judges, the alternative explanation that bias may be influencing results must be strongly considered. Certainly, the few attorney comments publicly reported from the "Judging the Judges" poll show that stereotyping is at play in this bar poll. The *Las Vegas Review-Journal* itself acknowledged the possibility that respondents might be evaluating women judges more harshly when it reported one attorney com-

menting that “[t]he female judges [in Family Court] are too emotional and biased, and they try too hard to make everyone happy” (Whitely 2006). This attorney is likely being influenced by stereotypes about women, namely that women are too emotional and nonconfrontational to be competent judges.

Women judges at every level report that their authority in the courtroom is challenged in ways that do not mirror what happens to their male colleagues. Should they fail to take control of their courtroom, they are perceived as weak and indecisive—or even incompetent. If they assert their authority, though, they will be labeled as discourteous or unduly punitive. In either case, their evaluations by members of the bar are negatively affected. These concerns are heightened further for minority women judges, who have faced even more intense scrutiny and challenges to their competence (Burger 2007).²³

There are reasons to think that the results of the “Judging the Judges” bar poll are not entirely aberrational. As Table 1 shows, judicial survey instruments used across the country are variations of the same ABA blueprint. Accordingly, the “Judging the Judges” survey closely tracks the bar surveys of other jurisdictions. An important area of further inquiry is to determine whether attorney surveys in other jurisdictions reflect the same unexplained gender and race biases. It is possible that the magnitude of unexplained bias is greater or lesser in other jurisdictions. It will be critical to identify these differences in comparative perspective so that we can isolate the characteristics of these programs that tend to minimize or exacerbate gender and race bias.

While unconscious bias cannot be entirely eliminated, questions could be asked in a way that minimizes the problem. These alternative formulations can be identified through careful comparative research. Experimental research may also help to develop a more comprehensive set of best practice recommendations for attorney survey instruments. Research shows that behaviorally oriented questions, which solicit information based on experiences, behaviors, and skills, may yield different responses than do attitudinal questions, which tend to be less specific (Dillman 2000; Schaeffer & Presser 2003).²⁴ As Missouri’s experience illustrates, however, question wording is not a silver bullet for eliminating the effects of unconscious bias (Burger 2007). By comparing the gender and race gaps in differently structured surveys, we can determine the

²³ We have not tested this interaction term here as, unfortunately, minority women judges are a null set in Nevada.

²⁴ For example, on the topic of timeliness, a behavioral question might read, “In your appearances before Judge X in the past year, how many times has he or she been late to court?” An attitudinal question may present the following statement, “Judge X is routinely late to court,” followed by a scaled answer set.

degree to which various alternative constructions may mitigate the problem of unconscious bias in judicial performance evaluations.

Conclusion

The findings of possible unconscious bias in the attorney bar poll raise significant questions about whether JPEs, as they are currently instituted, are fair. Certainly, the evidence is not good for the *LVRJ*'s privately run JPE. A poll that consistently scores women and minority judges lower, without regard to other observable measures of judicial quality, has an undesirable impact on democracy. Women and minority candidates will be more vulnerable to electoral challenges.

As there are increasing calls for reliance on JPEs as a way of ensuring quality standards, it is imperative that these processes not reproduce—even inadvertently—a system that disfavors groups like women and minorities, who have been historically underrepresented in the judiciary. Retention election outcomes have been linked to performance scores on judicial evaluations. Brody finds a significant linear relationship between JPE scores and voting outcomes, such that the higher the score a judge received on the JPE, the higher the voting percentage in favor of retention (Brody 2008). While similar research has yet to be completed on the specific electoral effects of the *LVRJ* “Judging the Judges” poll, it is reasonable to expect that a systematic downward bias in scores would harm the chances of women and minorities at the polls.

Unfair and biased evaluations do not only harm the individuals subject to them, but would have far-reaching and deleterious effects on the judiciary as an institution. Voters who have access to state-sponsored judicial performance evaluations rely on that information when casting their votes (Esterling 1998). If women and minorities are unfairly rated lower, this disparity will reproduce itself in the judicial election returns, compromising the diversity of our state benches. For this reason, it is fundamentally important for the democratic process of judicial elections that JPEs refrain from providing biased, uneven, or unfair ratings of judicial candidates.

References

- American Bar Association (1985) “Guidelines for the Evaluation of Judicial Performance.” Washington, DC: American Bar Association.
- (1992) *Legal Education and Professional Development: An Educational Continuum*. Washington, DC: American Bar Association.

- (2005) "Guidelines for the Evaluation of Judicial Performance with Commentary." Chicago, IL: American Bar Association.
- (2007) *ABA Model Code of Judicial Conduct*. Washington, DC: American Bar Association.
- American Psychological Association (1991) "In the Supreme Court of the United States: *Price Waterhouse v. Ann B. Hopkins*: Amicus curiae brief for The American Psychological Association," 46 *American Psychologist* 1061–70.
- Andersen, Seth S. (2000) "Judicial Retention Evaluation Programs," 34 *Loyola L.A. Law Rev.* 1375–89.
- American Tort Reform Foundation (2008) "Judicial Hellholes 2008/2009." Washington, DC: American Tort Reform Foundation.
- Baum, Lawrence (2001) *American Courts: Process and Policy*, 5th ed. Boston: Houghton Mifflin.
- Beck, Nathaniel, & Jonathan Katz (1995) "What To Do (and Not To Do) With Time-Series Cross-Section Data," 89 *American Political Science Rev.* 634–47.
- Bertrand, Maria, & Sendhil Mullainathan (2004) "Are Emily and Brendan More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination," 94 *American Economic Rev.* 991–1014.
- Bielby, William T. (2000) "Minimizing Workplace Gender and Racial Bias," 29 *Contemporary Sociology* 120–9.
- Bonneau, Chris W., & Melinda Gann Hall (2009) *In Defense of Judicial Elections*. New York: Routledge.
- Brenner, Saul, & Timothy M. Hagle (1996) "Opinion Writing and Acclimation Effects," 18 *Political Behavior* 235–61.
- Brody, David C. (2000) "Judicial Performance Evaluations by State Governments: Informing the Public While Avoiding the Pitfalls," 21 *Justice System J.* 333–56.
- (2008) "The Use of Judicial Performance Evaluation to Enhance Judicial Accountability, Judicial Independence, and Public Trust," 86 *Denver Univ. Law Rev.* 1–42.
- Burger, Gary K. (2007) "Attorney's Ratings of Judges: 1998–2006." Mound City, MO: Report to the Mound City Bar.
- Choi, Stephen J., et al. (2007) "Professionals or Politicians: The Uncertain Empirical Case for an Elected Rather than Appointed Judiciary," *Olin Working Paper*, University of Chicago Law and Economics, Chicago, IL.
- (2009) "Are Judges Overpaid? A Skeptical Response to the Judicial Salary Debate." 1 *Journal of Legal Analysis* 47–117.
- Cohen, Thomas (2006) "Appeals from General Civil Trials in 46 Large Counties, 2001–2005." Washington, DC: Bureau of Justice Statistics, U.S. Department of Justice.
- Colorado Office of Judicial Performance Evaluation (2008) "Judicial Performance Reviews." Available online at <http://www.coloradojudicialperformance.gov/review.cfm>. Accessed March 17, 2010.
- Commission on Women in the Profession (2009) "A Current Glance at Women in the Law 2009." Chicago, IL: American Bar Association.
- Committee on Judicial Independence (2006) "Resource Kit on Fair and Impartial Courts," Chicago, IL: American Bar Association.
- Cordell, La Doris H. (2008) "The Joy of Judging," 43 *Harvard Civil Rights—Civil Liberties Law Rev.* 639–43.
- Correll, Joshua, et al. (2007) "Across the Thin Blue Line: Police Officers and Racial Bias in the Decision to Shoot." 96 *J. of Personality and Social Psychology* 1006–23.
- Dillman, Don (2000) "The Role of Behavioral Survey Methodologists in National Statistical Agencies." 68 *International Statistical Rev.* 200–13.
- Dovido, John F., & Samuel L. Gaertner (2000) "Aversive Racism in Selection Decisions: 1989 and 1999." 11 *Psychological Science* 315–19.
- Downey, Nancy (2008) "Las Vegas Review Journal Judicial Performance Evaluation 2008." Las Vegas, NV: Downey Research Associates.

- Driscoll, John C., & Aart C. Kraay (1998) "Consistent Covariance Matrix Estimation With Spatially Dependent Panel Data," 80 *Rev. of Economics and Statistics* 549–60.
- Durham, Christine M. (2000) "Gender and Professional Identity: Unexplored Issues in Judicial Performance Evaluation," 39 *Judges' J.* 13–16.
- Epstein, Lee, et al. (2003) "The Norm of Prior Judicial Experience and Its Consequences for Career Diversity on the U.S. Supreme Court," 91 *California Law Rev.* 903–65.
- Esterling, Kevin M. (1998) "Judicial Accountability the Right Way," 82 *Judicature* 206–15.
- Esterling, Kevin M., & Kathleen M. Sampson (1998) *Judicial Retention Evaluation Programs in Four States: A Report with Recommendations*. Chicago, IL: American Judicature Society.
- Feeney, Floyd (1987) "Evaluating Trial Court Performance," 12 *Justice System J.* 148–70.
- Flango, Victor Eugene, & Craig R. Ducat (1979) "What Difference Does Method of Judicial Selection Make? Selection Procedures in State Courts of Last Resort," 5 *Justice System J.* 25–45.
- Glick, Henry R. (1978) "The Promise and the Performance of the Missouri Plan: Judicial Selection in the Fifty States," 32 *Univ. of Miami Law Rev.* 509–43.
- Glick, Henry R., & Craig F. Emmert (1986) "Selection Systems and Judicial Characteristics: The Recruitment of State Supreme Court Judges," 70 *Judicature* 228–35.
- Griffin, Jacqueline R. (1994) "Judging the Judges," 21 *Litigation* 5.
- Haire, Susan B. (2001) "Rating the Ratings of the American Bar Association Standing Committee on Federal Judiciary," 22 *Justice System J.* 1–18.
- Hedlund, Ronald D., et al. (1979) "The Electability of Women Candidates: The Effects of Sex Role Stereotypes," 41 *The J. of Politics* 513–24.
- Heilman, Madeline E. (1983) "Sex Bias in Work Settings: The Lack of Fit Model," 5 *Research in Organizational Behavior* 269–98.
- Heilman, Madeline E., et al. (2004) "Penalties for Success: Reactions to Women Who Succeed at Male Gender-Typed Tasks," 89 *J. of Applied Psychology* 416–27.
- Hettinger, Virginia A., et al. (2003) "Acclimation Effects and Separate Opinion Writing on the United States Courts of Appeals," 84 *Social Science Q.* 792–810.
- Hoechle, Daniel (2007) "Robust Standard Errors for Panel Regressions with Cross-Sectional Dependence," 7 *Stata Journal* 281–312.
- Hopkins, A. D. (2008) "Judging the Judges: Nearly 800 Lawyers Turn in Judicial Performance Evaluation Grades," *Las Vegas Rev. J.* May 18.
- Institute for the Advancement of the American Legal System (2008) "The Bench Speaks on Judicial Performance Evaluation: A Survey of Colorado Judges." Denver, CO: Institute for the Advancement of the American Legal System.
- (2010) "Current Projects: Judicial Performance Evaluation," Institute for the Advancement of the American Legal System. Available online at <http://www.du.edu/legalinstitute>. Accessed August 11, 2010.
- Jackson, Jeffrey (2007) "Beyond Quality: First Principles in Judicial Selection and their Application to a Commission-Based Selection System," 34 *Fordham Urban Law J.* 125–61.
- Kagan, Elena (2006) "Women and the Legal Profession—A Status Report," 61 *The Record* 37–48.
- Kearney, Richard C. (1999) "Judicial Performance Evaluation in the States," 22 *Public Administration Q.* 468–89.
- Kearney, Richard C., & Holly Sellers (1996) "Sex on the Docket: Reports of State Task Forces on Gender Bias," 56 *Public Administration Rev.* 587–92.
- Lane, Kristin A., et al. (2007) "Implicit Social Cognition and the Law," 19 *Annual Rev. of Law and Social Science* 1–25.
- Laube, Heather, et al. (2007) "The Impact of Gender on the Evaluation of Teaching: What We Know and What We Can Do," 19 *National Women's Studies Association J.* 87–104.

- Lauck, Scott (2006) "Missouri Bar Poll Has Little Pull with Voters," *Missouri Lawyers Weekly* 13 November.
- Martell, Richard F. (1991) "Sex Bias at Work: The Effects of Attentional and Memory Demands on Performance Ratings of Men and Women," 21 *J. of Applied Social Psychology* 1939–60.
- McMurdo, Doug (2010) "Judging the Judges: Many Lawyers Seek Appointment System," *Las Vegas Review Journal* May 14.
- Missouri Bar Judicial Evaluation Survey Committee (2007) "Report of the Judicial Evaluation Survey Committee." Jefferson City, MO: The Missouri Bar.
- Nieva, Veronica F., & Barbara A. Gutek (1980) "Sex Effects on Evaluation," 5 *Academy of Management Rev.* 267–76.
- O'Connor, Sandra Day (2009) "Judicial Independence and Civic Education," 22 *Utah Bar J.* 10–19.
- Pelander, A. John (1998) "Judicial Performance Review in Arizona: Goals, Practical Effects and Concerns," 30 *Arizona State Law J.* 643.
- Posner, Richard A. (2000) "Is the Ninth Circuit Too Large? A Statistical Study of Judicial Quality," 29 *The J. of Legal Studies* 711–19.
- (2005) "Judicial Behavior and Performance: An Economic Approach," 32 *Florida State Univ. Law Rev.* 1259–79.
- Rachlinski, Jeffrey J., et al. (2008) "Does Unconscious Racial Bias Affect Trial Judges?," 84 *Notre Dame Law Rev.* 1195–246.
- Rottman, David B. (2004) "Trends and Issues in the State Courts: Challenges and Achievements," in *The Book of the States: 2004*. Lexington, KY: Council of State Governments.
- Sackett, Paul R., et al. (1991) "Tokenism in Performance Evaluation: The Effects of Work Group Representation on Male-Female and White-Black Differences in Performance Ratings," 76 *J. of Applied Psychology* 263–7.
- Schaeffer, Nora Cate, & Stanley Presser (2003) "The Science of Asking Questions," 29 *Annual Rev. of Sociology* 65–88.
- Schafran, Lynn Hecht, & Norma J. Wikler (2001) "Gender Fairness in the Courts: Action in the New Millennium." Washington, DC: National Judicial Education Program, NOW Legal Defense and Education Fund.
- Slotnick, Elliot E. (1983) "The ABA Standing Committee on Federal Judiciary: A Contemporary Assessment—Part 1," 66 *Judicature* 349–62.
- Sprague, Joey, & Kelley Massoni (2005) "Student Evaluations and Gendered Expectations: What We Can't Count Can Hurt Us," 53 *Sex Roles* 779–93.
- Standing Committee on Judicial Independence (2000) "Standards on State Judicial Selection: Report of the Commission on State Judicial Selection Standards." Chicago, IL: American Bar Association.
- Steinpreis, Rhea E., Katie A. Anders, & Dawn Ritzke (1999) "The Impact of Gender on the Review of the Curricula Vitae of Job Applicants," 41 *Sex Roles* 509–28.
- Sternlight, Jean (1996) "Symbiotic Legal Theory and Legal Practice: Advocating a Common Sense Jurisprudence of Law and Practical Applications," 50 *Univ. of Miami Law Rev.* 707–78.
- Thomas, Rebecca M., et al. (2009) "Nevada Judicial Evaluation Pilot Project: Final Report." Reno, NV: Grant Sawyer Center for Justice Studies.
- Titus, Ronald R., et al. (2009) "Annual Report of the Nevada Judiciary, Fiscal Year 2009," Carson City, NV: Administrative Office of the Courts.
- Valian, Virginia (1998) *Why So Slow? The Advancement of Women*. Boston, MA: MIT Press.
- Walker, Thomas G., & Deborah J. Barrow (1985) "The Diversification of the Federal Bench: Policy and Process Ramifications," 47 *Journal of Politics* 596–617.
- Whitely, Joan (2006) "Judging the Judges: About the Survey," *Las Vegas Review-Journal* Apr. 30.

Cases Cited

Caperton v. A. T. Massey Coal Co., Inc. (2009) 556 U.S. ____.

Price Waterhouse v. Hopkins (1989) 490 U.S. 288.

Rebecca D. Gill is an Assistant Professor of Political Science at the University of Nevada Las Vegas. Her research interests include comparative judicial institutions and American judicial selection.

Sylvia R. Lazos is the Justice Myron Leavitt Professor of Law at the William S. Boyd School of Law, University of Nevada Las Vegas. Her research interests include judicial diversity and critical race theory as applied to constitutional law.

Mallory M. Waters is a graduate student at the University of Nevada Las Vegas. Currently, she is working in the Special Projects Division of the San Diego Superior Court as part of the Capitol Fellows Program in Judicial Administration.