Beyond accountability and independence
Judicial selection and state court performance

by DAMON CANN

Political scientist Douglas Rae has shown that the method by which governmental officials are selected can have profound political consequences. It thus behooves judicial scholars to evaluate the quality of justice obtained under different selection plans. Not only could the results of such studies have important ramifications for states, they can also inform the nascent debate on whether federal judges should be elected rather than appointed.

Political pressures on judicial candidates led many states to abandon judicial elections in favor of merit selection plans (among which the Missouri Plan is the most prominent) where judges are appointed to an initial term of office by the governor (who chooses from a list prepared by a nonpartisan nominating commission). After these judges serve their initial term, they must stand for a retention election. A host of “good government” groups, as well as the American Judicature Society, have relentlessly advocated reforming state courts by adopting such merit-selection plans nationwide.

Proponents of merit selection contend that the merit-based system is an ideal compromise between two conflicting ideals for judicial selection plans—independence and accountability. Independence is protected because judges are initially appointed, generally serve longer terms, raise and spend little money on campaigns, and do not face other candidates when they stand for retention elections (and thus campaigns tend to focus on qualifications rather than ideology). Still, the retention elections allow for a measure of accountability by providing a means for the public to remove a judge whose decisions are unacceptable.

Some have contended that the popularity of merit selection is unwarranted because limited empirical evidence exists supporting the claim that merit selection actually preserves independence while promoting accountability. I contend that the scholarly obsession with independence and accountability has led scholars to overlook the overall quality of outcomes under different selection systems. This article reviews criticisms of common modes of selection and the empirical evidence behind those criticisms, and recommends a new way to evaluate the overall performance of a state court system based on the evaluations of the judges within the system. Judges in states with certain modes of selection may rate the overall quality of justice in their state at a higher level than judges in states with other modes of selection.

Empirical studies
Perhaps the most striking factor in existing empirical studies of judicial selection mechanisms is the lack of support for reformers claims. While reformers have contended that merit selection plans yield more qualified judges, the bulk of the evidence suggests this is not the case. For example, one study examining the personal

How would you classify the judiciary in your state?

How would you rate the job being done by courts and judges in your state?

Question wording

The dependent variable for this analysis measures the quality of justice in a state using a general performance measure. The question reads, "How would you rate the job being done by courts and judges in your state?" The response set is Excellent, Good, Fair, Poor, and Don't Know.

The variable male is based on the questions, "What is your gender?" The response set is Male and Female.

The race variables (African-American, Hispanic, and Other) are constructed from a single question, "What is your race?" The response set is White, Black, Hispanic/Latino, and Other.

The variables regarding the extremity of the state judiciary are created from a single question, "How would you classify the judiciary in your state?" The response set is Very Liberal, Somewhat Liberal, Moderate, Somewhat Conservative, Very Conservative, and Don't Know. The variable Very Extreme is coded 1 for judges who responded Very Liberal or Very Conservative; the variable Somewhat Extreme is coded 1 for judges who responded Somewhat Liberal or Somewhat Conservative.

The variable for respondent age is based on the question, "In what year were you born." The variable received the value of the judges year of birth subtracted from 2002.

The mode of selection for most judges in the state comes from Laura Langer’s web site at www.u.arizona.edu/~langer/NSFESLANGER.htm.
It is equally plausible that contributors decide to donate to judges who already support their views (rather than giving money to someone they disagree with in hopes that their contribution will change a judge’s mind). The latter possibility is in harmony with studies of the link between campaign contributions and votes in Congress that show that the correlation between votes and money is best understood as ideological groups supporting candidates who are already disposed to vote with the group rather than aggressive donors trying to buy the votes of legislators. In short, available evidence does not show that campaign fundraising leads to a general decline in the integrity of judges.

Elections have also been faulted for failing to promote accountability—the very virtue they are purported to provide. Indeed, judicial elections have been characterized as low-profile campaigns with low rates of participation and little discussion of issues. However, more recent evidence challenges this notion. A study of Ohio voters finds that citizens are able to look judicial candidates to a political party and use that information in casting their votes, even though party affiliation does not appear on the ballot in Ohio judicial elections. The study also shows that ideology influences voting behavior in judicial elections. Further, voters seem to have particularly high levels of participation and understanding when voter guides and other independent sources of campaign information are available. Additionally, a landmark multi-state study shows that voters in elections (notably including retention elections) are aware of issues in elections and that those issues influence electoral outcomes. Whatever judicial elections may have been in the past, the bulk of empirical evidence suggests that the key ingredients for accountability are present in contemporary judicial elections.

The general interpretation of these studies has been that elections may be no worse than merit selection in terms of accountability and independence. However, the actual findings may be more concisely summarized by saying that evidence shows that elections do, in fact, promote accountability and that retention elections do not provide as much insulation from external political forces as had been hoped.

Regardless of the interpretation, the problem faced by the literature on accountability and independence is a general failure to link the results to conclusions about overall court performance. After all, the reason we are concerned about accountability and independence is that we think systems with accountable and/or independent judges will yield better overall outcomes. This being the case, explorations into the value of the sundry judicial selection systems would be better carried out by directly examining the quality of outcomes rather than examining independence, accountability, or other intermediary factors leading to quality overall outcomes.

8. For a comprehensive discussion, see John R. Wright, Interest Groups and Congress: Lobbying, Contributions, and Influence, (Boston, MA: Allyn and Bacon, 1996).
9. For examples, see Patrick W. Dunn Judicial Selection and the States: A Critical Study with Proposals for Reform, 1 Hofstra L. Rev. 283 and Philip Dubois, From Ballot to Bench: Judicial Elections and the Quest for Accountability (Austin, TX: University of Texas Press, 1980).
12. Hall, supra n. 5.
View from the bench

Any empirical study of the overall quality of court outcomes runs into one major obstacle: Defining and measuring the quality of a state court system outcomes. A unique, nationwide survey of 2,428 state court judges conducted by the Justice at Stake Campaign provides some traction on this difficult issue. The sample includes judges from state supreme, appellate, and trial courts. (See page 227 for the wording of survey questions).

The survey asked judges to list the most important duties of courts and judges. The four responsibilities...
court (the most salient elections in a state) produces similar results. Thus, while this operationalization of mode of selection may be imperfect, the results are robust to several different operationalizations.15

Additionally, state courts with more resources may be able to better carry out their work. Thus, I include an index measuring the professionalism of a state's courts.16 While the items in this index bear most directly on state supreme courts (salary of supreme court judges, staff for the high court, and similar items), the professionalism of state high courts and lower courts is likely highly correlated, so the index at hand serves as a proxy for the professionalism of the entire system.

The characteristics of the judge responding to the survey may also have bearing on his or her evaluations of the quality of justice. To control for the possible effect of race, I include a set of dichotomous indicators for African-American, Hispanic, and Other Non-white; white is the baseline category. In addition to the respondent's race, the respondent's gender may also affect their views of state court performance. Accordingly, the variable Male is coded 1 for male judges and 0 for females.

Judges' length of service on the bench may affect their attitudes toward the quality of their state courts. While the survey did not directly ask respondents how long they had served on the bench, it did ask their age. Age should be correlated with length of service on the bench (though not perfectly). Moreover, age of itself may affect judges' attitudes toward their state courts. Accordingly, I include a variable for age in years at the time of the survey.

One final characteristic of the judges must be considered—judges serving on higher level courts may have different attitudes about their state courts than judges serving on lower level courts. Accordingly, I include a variable for respondents who serve on their state supreme court and state appellate courts, with trial courts being the baseline category.

Finally, a judge's views of other judges in the state may affect the way he or she evaluates how well state courts and judges perform. As such, I include a measure of the respondent's assessment of the ideological extremity of other judges in the state. Very Extreme is coded 1 for judges who think judges in their state are either very liberal or very conservative. Somewhat Extreme is coded 1 for judges who think judges in their state are either somewhat liberal or somewhat conservative. Judges who think their state has a moderate judiciary constitute the baseline category.

Results
Because the dependent variable consists of four ordered categories (excellent, good, fair, and poor), I use an ordered probit model for the statistical analysis.17 Some judges did not respond to several questions, leaving us with a usable n of 2,230. The key result is that judges in states where most judges are elected by merit selection or appointment rate their state court system significantly higher than judges in states where most judges are elected in partisan elections. The results also show that judges' evaluations in nonpartisan states do not differ significantly from the evaluations of judges in partisan states. In short, the data show that judges selected in traditional elections (whether partisan or nonpartisan) rate their state court systems lower than judges in states that use a merit selection or an appointment system.

Because the model with the full typology of selection systems provides more detailed information, our interpretation of results will focus on that model, though the results from both models were substantively similar.18 Figure 1 shows the relationship between mode of selection and judges' ratings of their state court systems graphically using predicted probabilities.18 These figures represent the probability of a judge with specific characteristics rating their system at each of the 4 levels—excellent, good, fair or poor. The difference between partisan and nonpartisan systems is negligible. However, judges in merit selection states have a 37 percent chance of rating their state courts as excellent compared to just a 32 percent chance in partisan states. Remarkably, judges in states where judges are appointed by the governor or legislature have a probability of about .5 of rating their state courts as excellent. As the predicted probability of being in the "excellent" category increases (moving from partisan toward appointment plans), the probability of being in the "good," "fair," or "poor" categories decreases. While there is substantial variation across systems in judges evaluations of performance, it is worth noting that it is very unlikely for judges in any system to rate their state courts as "poor."

The court professionalism index has no significant impact on how judges rate their state courts. Two possible explanations exist. First, when evaluating the quality of justice in their states, judges may evaluate the extent to which judges do the best they can given available resources. If evaluations are indeed made in this way, the resources available to judges would not likely affect the ratings. Second, it may be that if the index captured the professionalism of lower courts as well as it assesses state supreme courts then one would find a significant effect.

The characteristics of the judges

15. A third possible problem is that many states fill mid-term court vacancies through appointment even in elective states. Still, these individuals must ultimately face election, which could still affect judges' evaluations of how well their court system functions.


17. A table containing the statistical results is available at http://dcanin.myweb.uga.edu/judges.pdf.

18. Further, a likelihood ratio test shows that this model fits the data significantly better than the model with the elected/appointed dummy.

19. In an ordered probit model, the effects of a variable are conditional upon the values of all other variables in the model. To generate these predicted probabilities, I varied the mode of selection while setting state court professionalism and age to their respective means and assumed a state judiciary that is somewhat extreme (the modal category), and a respondent who is a white male trial court judge (again reflecting the modal categories).
answering the survey also affect the way they evaluate the quality of justice in their state. Relative to white judges, African-American judges rate the performance of their state courts lower. However, Hispanic judges and judges of other races do not differ significantly from white judges in their evaluation of state court performance. Not only does race affect judges’ evaluations of the quality of justice in their state courts, but gender does as well. Male judges tend to rate the performance of their state courts higher than female judges. Age and the level of court a judge serves on do not have significant effects on judges’ evaluations of the quality of their state courts.

Finally, a judge’s assessment of the extremity of other judges in their state court system has a strong, significant effect on their evaluations of their state courts. Judges who view the other judges in their state as too extreme rate the performance of their state courts much lower. Figure 2 graphs the predicted probabilities of rating state court performance varying judges’ evaluations of the extremity of judges in their state, holding the method of selection constant as merit selection. Judges who consider their state judiciary to be ideologically moderate have a .41 probability of rating their state courts as excellent. In contrast, for judges who see their state judiciary as somewhat extreme or very extreme, the probability of an “excellent” rating drops to .37 or .26, respectively. It appears that a judge’s assessment of the ideology of judges in his or her state affects their assessment of the quality of justice about as strongly as the mode of selection.

Discussion
While some have criticized reformers for advancing their cause without empirical evidence of its effectiveness, this analysis provides strong empirical evidence that merit selection and appointment plans are superior to partisan and nonpartisan elections. In states where most judges are chosen by merit selection or appointment, judges rate the quality of the work done by their state courts higher than judges in states where most judges are selected through partisan or nonpartisan elections.

We began with a discussion of the
two issues that have characterized the debate over judicial selection mechanisms—accountability and independence—and noted that empirical research has supported the notion that elections maximize accountability. However, even if citizens are well-informed and able to use elections to maximize accountability, the evidence here shows that, on balance, using elections to select judges actually diminishes the overall performance of state courts (as measured in the two ways considered in this article). It may be that accountability is not the key to an effective judiciary. Alternatively, while accountability could be desirable, it may be impossible to obtain it without the negative side-effects inherent to elections.

More research is necessary to target the specific reasons why judges in merit selection and appointment states rate their courts higher than judges in election states. One possibility is that judges are concerned that the politicization of judicial campaigns may be damaging the legitimacy of the judiciary in their state. Scholars should consider how the mode of selection effects citizen perceptions of the courts. Another possibility lies in campaign finance. Judges may view fundraising and campaigning as a tremendous burden. Further, while there is scant evidence that campaign contributions effectively “buy” judges’ votes, it is possible that the very appearance of quid pro quo exchanges makes people uncomfortable.

An additional avenue for future exploration is the measurement of the quality of justice. While the measure used here based on judges’ evaluations of their state court systems has strong face validity, future work should consider alternative measures of the general quality of court outcomes. This will allow researchers to test the robustness of these results.

Ultimately, this research provides solid empirical evidence that the campaign to move toward merit selection systems and away from election systems is both warranted and desirable. While judicial elections may not be uncompetitive events with uninformed voters, merit selection and appointment states appear to achieve better overall court performance. Reformers should continue to press for their cause knowing that the general outcomes of court systems using merit selection or appointment plans are superior to systems using partisan or non-partisan elections. 22

DAMON CANN is an assistant professor in the Department of Political Science at the University of Georgia. (dcann@uga.edu)