

THE CONTINUING SEARCH FOR A MEANINGFUL MODEL OF JUDICIAL RANKINGS AND WHY IT (UNFORTUNATELY) MATTERS

SCOTT BAKER†

ADAM FEIBELMAN††

WILLIAM P. MARSHALL†††

INTRODUCTION

Judicial Evaluations and Information Forcing: Ranking State High Courts and Their Judges (Judicial Evaluations),¹ by Professors Stephen Choi, Mitu Gulati, and Eric Posner is an important effort to create and apply an objective methodology for evaluating the relative quality of state supreme courts. The article follows a previous work by Professors Choi and Gulati, *Choosing the Next Supreme Court Justice: An Empirical Ranking of Judge Performance (Choosing the Next Supreme Court Justice)*,² which employed a similar methodology to evaluate the quality of federal appeals court judges.³

Copyright © 2009 by Scott Baker, Adam Feibelman, and William P. Marshall.

† Professor of Law, UNC Chapel Hill, School of Law.

†† Associate Professor Law, UNC Chapel Hill, School of Law.

††† Kenan Professor Law, UNC Chapel Hill, School of Law.

1. Stephen J. Choi, Mitu Gulati & Eric A. Posner, *Judicial Evaluations and Information Forcing: Ranking State High Courts and Their Judges*, 58 DUKE L. J. 1313 (2009).

2. Stephen J. Choi & G. Mitu Gulati, *Choosing the Next Supreme Court Justice: An Empirical Ranking of Judge Performance*, 78 S. CAL. L. REV. 23 (2004) [hereinafter Choi & Gulati, *Choosing the Next*]; see also Stephen Choi & Mitu Gulati, *A Tournament of Judges?*, 92 CAL. L. REV. 299, 299–304 (2004) [hereinafter Choi & Gulati, *A Tournament of Judges?*] (proposing a system to evaluate federal judges).

3. For additional perspectives on judicial performance by Professors Choi, Gulati, and Posner, see generally Stephen J. Choi & G. Mitu Gulati, *Mr. Justice Posner? Unpacking the Statistics*, 61 N.Y.U. ANN. SURV. AM. L. 19 (2005) (examining the background reasons for Judge Posner's high performance on rankings); Stephen J. Choi & G. Mitu Gulati, Essay, *Ranking Judges According to Citation Bias (As a Means to Reduce Bias)*, 82 NOTRE DAME L. REV. 1279

Professors Choi, Gulati, and Posner consistently advance the theme that judicial rankings are here to stay.⁴ They contend that it is important to develop an objectively neutral rankings system for two reasons. First, an objective rankings system will offer significant insights into judicial performance precisely because it is transparent and replicable. Second, an objectively neutral methodology will counter other evaluation systems that are more designed to reflect the agendas of the evaluators than the quality of the judges studied.

We *generally* agree with Professors Choi, Gulati, and Posner with respect to these central assertions. Judicial rankings, for better or worse, are not going away. Therefore, there is a strong need to develop meaningful and appropriate measures for evaluation. We also agree with the authors that developing objective criteria, if possible, can unmask and refute the efforts of interest groups and politicians to manipulate standards to further their own agendas.

The Choi, Gulati, and Posner approach is straightforward: (1) identify attributes of judicial quality that are value neutral and uncontestable, (2) find objectively measurable proxies that correlate with those attributes, and (3) rank state courts according to the proxies. The judicial attributes Professors Choi, Gulati, and Posner select are productivity, influence, and independence. The measurable proxies are the number of opinions written by state courts, the amount of out-of-state citations to those opinions, and instances of disagreement among judges of the same political stripe. The study suggests that to the extent that these proxies measure uncontestable attributes of good judges and good courts, the rankings should reflect relative judicial quality. Equally important, the rankings should motivate judges and courts to improve their ranking by improving their performance with respect to the underlying attributes.

That said, we have a number of concerns with the authors' approach. To initially illustrate these concerns, consider the following example. Suppose one wants to rank seafood restaurants. Convenience is at least one important attribute in a restaurant, and one possible proxy for convenience might be the distance of the seafood place from the closest highway exit ramp. One might

(2007) (proposing a ranking system to make judicial bias more transparent); Stephen J. Choi & G. Mitu Gulati, Essay, *Which Judges Write Their Opinions (And Should We Care?)*, 32 FLA. ST. U. L. REV. 1077, 1079–96 (2005) (discussing the desirability of addressing issues of authorship and delegation in federal court opinions).

4. See, e.g., Choi et al., *supra* note 1, at 1364.

therefore develop a restaurant quality index based on the ramp-to-entrance distance in which restaurants closer to a ramp are ranked higher than those farther away. But how effective would this ranking actually be in determining restaurant quality?

First, the ranking on its own provides little meaningful information if convenience is much less important to customers than other attributes. It tells nothing, for example, about the quality of the food, the friendliness of the service, or the cleanliness of the facility. Second, if the proxy—ramp-to-entrance—is a small part of what it means to be convenient, the ranking does not even tell one much about convenience. Customers, after all, might find larger parking lots to be of more convenience than highway proximity. Finally, because a ranking encourages its subjects to take actions to increase their scores, this convenience ranking might lead a restaurant to move closer to a highway rather than take more important actions such as improving the quality of its food, hiring friendlier employees, working on its cleanliness, or increasing the size of its parking lot.

These same concerns come into play in the Choi, Gulati, and Posner study. Similar to the choice of convenience in the restaurant ranking described above, we strongly suspect that the attributes that the authors select in *Judicial Rankings* constitute relatively minor aspects of judicial quality. Those who care about state supreme courts may not be all that concerned about productivity, influence, or independence. Rather, observers may care more about whether judges have a particular ideology or judicial philosophy, whether they vote certain ways in hot button cases, whether they are fair, whether they present their legal opinions clearly and address opposing arguments, whether they possess a sound judicial temperament, whether they have unquestioned integrity, whether they treat lawyers respectfully, and whether they have outstanding reputations as jurists in their communities. These factors (with the exception of ideology and voting patterns) are undoubtedly hard, perhaps impossible, to quantify. And many of these factors, unlike the ones relied on by Professors Choi, Gulati, and Posner, cannot be culled from Westlaw or Lexis. Nonetheless, in deciding what is a high-quality court, the Choi, Gulati, and Posner study—and the presumption it generates—elevates measurable over nonmeasurable attributes without regard to their relative importance.⁵

5. The authors' response to these concerns is that the presumptive marker of judicial quality created by the ranking encourages parties to convey other useful information in an effort

Second, even if judicial productivity, influence, and independence are central to evaluating judicial quality, it is not at all clear that the authors' methodology accurately captures performance in these areas. That is to say, the correlation between the proxy and the attribute might not be that high. Does the number of published opinions, for example, really reflect judicial productivity? Or are there other more important factors involved, such as the total number of cases reviewed, the time a court takes between hearing and resolving a matter, or the effort the court devotes to judicial administration and supervision? If so, then, the *single* proxy they have chosen does not offer much more information about judicial productivity than distance from the highway offers about the convenience of restaurants in a world in which people care more about parking or hours of operation.

Third, in the same way that we are wary of a restaurant rankings system that only encourages restaurants to move closer to a highway, so are we concerned with the incentives for judicial behavior created by the Choi, Gulati, and Posner rankings. Evaluation systems encourage their subjects to engage in activity to improve their rankings.⁶ Indeed, rankings, as will be discussed, are beneficial when they encourage aspects of judicial behavior that should be rewarded and penalize aspects of judicial behavior that should be discouraged.⁷ But one must be confident in the correlation between the proxy and the attribute for the ranking to serve a salutary purpose. It is not clear that motivating judges to write more opinions, to work to have their opinions more frequently cited, or to disagree with judges of the same political party would increase the overall quality of state courts—particularly when those actions may require the judges to forego other activities that may be more important to the administration of justice.⁸ For example, judges striving to improve their productivity

to combat the presumption. But, as we will discuss *infra* Part II.B, that will be hard to do. A seafood restaurant, for example, grappling with a low convenience ranking would have to post flyers about the quality of the fish or the ambience of its dining room. These unverifiable claims about immeasurable quality attributes look self-serving and will be unlikely to ever dislodge the initial ranking presumption.

6. See, e.g., Frederick Schauer, *Incentives, Reputation, and the Inglorious Determinants of Judicial Behavior*, 68 U. CIN. L. REV. 615, 630 (2000) (discussing the role of reputation in influencing judicial behavior).

7. See *infra* Part II.C.

8. The authors readily concede this point. See Choi et al., *supra* note 1, at 1318 (citing Steven Kerr, *On the Folly of Rewarding A While Hoping for B*, 18 ACAD. MGMT. J. 769, 778 (1975)).

score by writing more opinions might shirk on other productive responsibilities such as court administration, careful consideration of the merits of each case, or a close reading of the trial record.

We therefore raise the possibility that the Choi, Gulati, and Posner study may serve to weaken rather than improve judicial performance. Again, we do not suggest that rankings systems are necessarily bad because they may prospectively affect judicial behavior. The key, however, is creating a metric that leads to beneficial results, and we are concerned that the particular criteria used in *Judicial Evaluations* are not so finely calibrated.

This Essay proceeds as follows. Part I addresses the threshold question of whether ranking state supreme courts is valuable. Assuming that there is value in the project, Part II looks at alternative measures for evaluating the performance of state courts, juxtaposing those measures with the indices of productivity, influence, and independence used by Professors Choi, Gulati, and Posner. Part III examines the Choi, Gulati, and Posner study in detail. It first questions whether, in creating a presumptive marker of judicial quality, the authors miss something by focusing on productivity, influence, and independence. Part III then discusses some of the methodological problems with the selected proxies. Part IV suggests an alternative lens of judicial evaluation that might lead to more beneficial results.

I. WHY RANK STATE SUPREME COURTS?

Prior to the Choi, Gulati, and Posner study, the most prominent rankings of state supreme courts were those issued by the United States Chamber of Commerce (Chamber).⁹ The Chamber surveys lawyers and asks them to evaluate the judicial impartiality, the judicial competence, and the overall quality of state courts. The rankings reward those courts that leading business lawyers perceive as friendly to business.¹⁰

The immediate positive result of *Judicial Evaluations* is that it provides a counterbalance to the Chamber by demonstrating that a high probusiness ranking does not equate to a high, objective quality ranking. This contribution is not to be undervalued. One use of any

9. See, e.g., HARRIS INTERACTIVE, INC., 2008 U.S. CHAMBER OF COMMERCE STATE LIABILITY SYSTEMS RANKING STUDY (2008), available at http://courts.delaware.gov/Courts/Superior%20Court/pdf/?harris_2008.pdf.

10. *Id.* at 20 tbl.7.

judicial rankings system is political spin. As Professors Choi and Gulati noted in their previous work, political actors consistently claim that the judges they support are of the highest caliber and quality,¹¹ and they will no doubt be quick to cite a ranking system that supports their claim if at all possible.¹² Accordingly, it is also unsurprising that rankings have also been used in debates over judicial salaries and resources.¹³ Indeed because of their “ease of transmission and understandability,”¹⁴ ranking systems have particularly powerful resonance in the public mind.¹⁵ For this reason alone, offering an objective alternative to the Chamber’s agenda-driven ranking system is invaluable.

Beyond providing a weapon in the spin wars, however, the benefits of ranking state supreme courts are not obvious.¹⁶ For example, although the project of ranking federal appeals court judges may make some sense as a method for determining the relative merit of particular judges competing for a United States Supreme Court nomination,¹⁷ the same rationale does not apply to comparing the relative merits of state supreme courts. After all, seats on the United States Supreme Court are filled by individuals, not full courts.¹⁸

Nor does ranking state supreme courts accomplish the same function as ranking institutions such as colleges, professional schools,

11. Choi & Gulati, *Choosing the Next*, *supra* note 2, at 39 n.29.

12. To be sure, politicians might shy away from using a ranking system that may be seen as too controversial or otherwise inconsistent with their agenda. A conservative, for example, would likely not be inclined to cite a high ranking from the American Civil Liberties Union as favorable evidence of a judge’s merit.

13. See Letter from Robert D. Ray, Iowa Judicial Comp. Task Force, to Nicholas Critelli, President, Iowa State Bar Ass’n (Apr. 18, 2005), available at [http://www.iowabar.org/misc/documents.nsf/2b85a4ea12f4bfac8625669d006e27ab/f0fb92e322a9987d86256ff20049a0bb/\\$FILE/Judicial%20compensation.pdf](http://www.iowabar.org/misc/documents.nsf/2b85a4ea12f4bfac8625669d006e27ab/f0fb92e322a9987d86256ff20049a0bb/$FILE/Judicial%20compensation.pdf) (cited in Choi et al., *supra* note 1, at 1316 n.2).

14. Choi & Gulati, *Choosing the Next*, *supra* note 2, at 39 n.29.

15. See, e.g., Press Release, Univ. of Iowa News Serv., Study by UI Sociologist Shows Law School Rankings Affect Student Decisions (Mar. 1, 2006), http://www.newsreleases.uiowa.edu/2006/march/030106law_study.html.

16. One reason for rankings suggested by the authors is to establish bragging rights. Choi et al., *supra* note 1, at 1316–17. Perhaps a ranking system could be justified on that count alone. But we are skeptical. The more likely uses of rankings are to (1) further policy or political agendas or (2) decide on the allocation of resources to the court.

17. See Choi & Gulati, *Choosing the Next*, *supra* note 2, at 25–32.

18. Rankings, of course, could be used to compare the relative merits of state supreme court judges but that would require individual rankings and not a ranking of the court as a whole. In this respect, it is notable that *Judicial Evaluation* does not rank all individual state court judges. See Choi et al., *supra* note 1, at 1360. Instead, it identifies judges that score particularly well on certain measures. See *id.* at 1361 tbl.14.

or hospitals.¹⁹ Those rankings influence the choices of prospective patients, students, and faculty to affiliate with a particular institution.²⁰ The only equivalent of “applicants” to state supreme courts are litigants. The aspect of judicial reputation of most interest to the litigant, however, is the court’s sympathies for, or against, their legal position—not whether that state’s supreme court is considered by some measures to be objectively excellent.²¹

Professors Choi, Gulati, and Posner suggest that ranking state courts may assist individuals or entities that have an interest in the influence of out-of-state courts. They claim that rankings may be useful to individuals or entities who are affected by out-of-state supreme court decisions because those parties then may have reason to try to affect judicial composition in the highly ranked (and therefore ostensibly more influential) states or they may want to submit amicus briefs to influence the development of the law in those states.²²

They further contend that lawyers, academics, and others who conduct legal research may be especially interested in knowing which court systems are seen as producing the best judicial opinions so they can determine the best examples of legal reasoning.²³ Finally, they might suggest in this same vein that rankings could be beneficial because a court in one state might follow a precedent from a highly rated out-of-state supreme court over a decision of a lesser-ranked state supreme court.

But these effects are likely to be marginal at best. The interests of individuals or entities in out-of-state proceedings are likely to be issue-dependent, not generic. For example, although labor law

19. See *Special Report: America’s Best Graduate Schools: Schools of Law*, U.S. NEWS & WORLD REP., Apr. 7–14, 2008, at 46, 46; *Special Report: America’s Best Graduate Schools: Schools of Medicine*, U.S. NEWS & WORLD REP., Apr. 7–14, 2008, at 36, 36–38; U.S. News & World Report, *America’s Best Hospitals*, <http://health.usnews.com/sections/health/best-hospitals/index.html> (last visited Apr. 4, 2009).

20. See, e.g., Michael Sauder & Ryon Lancaster, *Do Rankings Matter? The Effects of U.S. News & World Report Rankings on the Admissions Processes of Law Schools*, 40 LAW & SOC’Y REV. 105, 105 (2006) (finding that the *U.S. News & World Report* rankings “have significant effects on both the decisions of prospective students and the decisions schools make in the admissions process”).

21. For a discussion on the tendency of partisan elected state court judges to redistribute wealth from out-of-state defendants, see Eric Helland & Alexander Tabarrok, *The Effect of Electoral Institutions on Tort Awards*, 4 AM. L. & ECON. REV. 341, 341 (2002). “[D]ifferences in awards are caused by differences in electoral systems, not by differences in state law.” *Id.*

22. Choi et al., *supra* note 1, at 1317–18.

23. *Id.*

practitioners in one state may want to influence the development of labor law in another state, it is unlikely that they will have enough interest in that state's legal system to also seek to influence that state's law in other substantive areas. Likewise, legal researchers are likely to focus on cases in their substantive areas rather than on the reputations of out-of-state courts in general. They will normally investigate significant cases in their area of interest, regardless of jurisdiction, and will not artificially confine their inquiries to highly ranked judicial systems. Finally, a state supreme court seeking support for its holdings from another jurisdiction will be far more likely to be interested in the result of the case it cites than in the reputation of the court that decides it.²⁴ Indeed it is hard to believe that a court faced with conflicting decisions from two different courts would decide to follow one solely because the issuing court has a superior reputation.

There is one area, however, in which the impact of judicial rankings may be understated. Judicial rankings do not merely reflect past performance of a judge or a court system—they might also create incentives for a judge or court to take actions to improve their score.²⁵ If the ranking is used by the appointing authority—voters or governors—to grade court performance, scoring well on the ranking system will lead to job security and maybe even more resources.²⁶ And even if there are no tangible benefits to higher rankings, judges may seek to score better in a ranking system simply because they believe it will help improve their reputation and stature.²⁷ Accordingly, one reason to rank judges or courts would be to influence behavior. As such, the stakes in judicial rankings are high and the importance of getting it right and developing a system that serves to encourage appropriate judicial behavior is manifest. The

24. Contrast this with the reputation of the judge who wrote the cited case. Sometimes, an opinion will delineate which judge wrote the cited opinion to give the citation more authority. This is especially true if a famous and influential judge wrote the opinion. See David Klein & Darby Morrisroe, *The Prestige and Influence of Individual Judges on the U.S. Courts of Appeals*, 28 J. LEGAL STUD. 371, 371 (1999) (discussing the invocation of judges' names in citations of court opinions).

25. Professors Choi, Gulati, and Posner recognize this. Indeed, it is one reason they developed their ranking system. Choi & Gulati, *A Tournament of Judges?*, *supra* note 2, at 313–15.

26. In the competition for state dollars, courts might spin rankings a number of ways. A low-ranking court might attribute their rank or status to a lack of resources. A high-ranking court might claim that they deserve more money precisely because they are so excellent.

27. See Schauer, *supra* note 6, at 627–31.

question of whether the ranking system developed by Professors Choi, Gulati, and Posner actually correlates to the estimable qualities of a judge or a court is therefore of considerable consequence.

II. MEASURING JUDICIAL QUALITY

The question of what makes a good judge depends in no small part on who is asking. To the Chamber, the question is easy—a judge that lawyers perceive advance a probusiness agenda. Likewise, the Sierra Club presumably considers a good judge to be one whose opinions support environmental interests; organized labor considers a good judge to be one whose jurisprudence furthers workers' rights; and the American Medical Association considers a good judge to be one whose decisions limit tort awards for malpractice and otherwise demonstrate solicitude for the concerns of physicians and other medical professionals.

To a practicing attorney, the answer may be something different. No doubt any attorney would be delighted to appear before a judge who, for ideological reasons, is predisposed to rule in his favor. But we are not convinced that all lawyers are so obsessively self-interested that they would believe that simply because a judge might rule for that lawyer's side, that judge is thereby excellent. Moreover, in many, if not most, cases ideology does not come into play.²⁸ Thus, we suspect that practicing lawyers would most value judges who are impartial, clear, open-minded, fair, experienced, timely, and even-keeled in judicial temperament.²⁹

The academic or expert observer, in turn, may be more focused on qualities such as vision, creativity, writing quality, and scholarly impact. Concededly, academics or experts may be influenced in their appraisal by agreement with the judge on key issues, but they will also be attuned to the judge's knowledge of the law, ability to anticipate the implications of the judge's decisions, and intellectual honesty in addressing opposing arguments. When Michael W. McConnell, a prominent conservative, was nominated to the Tenth Circuit by President George W. Bush, for example, he was supported by over a

28. RICHARD A. POSNER, *HOW JUDGES THINK* 27–28 (2008).

29. See MODEL CODE OF JUDICIAL CONDUCT R. 2.2 cmt. 1 (2007) (calling for judges to serve with impartiality, fairness, and to be “objective and open-minded”); *id.* 2.5 cmts. 1–4 (instructing judges to perform duties competently and diligently; to be efficient, punctual, and prepared; and to avoid any “unnecessary cost or delay”).

hundred legal scholars, including some of the most liberal academics in the country, because of the strength of his scholarly contributions.³⁰

A president or governor, on the other hand, may view judicial quality in still another light. The president or governor wants to appoint a judge who shares a similar legal vision and who will use the judicial role to advance the appointing agent's political agenda.³¹ Yet, the president or governor will also be alert to other factors such as ethnic and geographic diversity and support from key constituencies or political actors. More to the point for our purposes, the president or governor is also likely, for both short-term and long-term reasons, to want to appoint a jurist with high stature, meaning a reputation for excellence among people of all political stripes. In the short term, the appointment of a highly regarded jurist will improve the president or the governor's own political capital as the choice will be seen as motivated by merit rather than politics. Further, establishing a pattern of appointing highly regarded jurists may make for a smoother confirmation of the president or governor's subsequent judicial nominees as an established track record of reputable appointments means that each additional appointment might be subject to less scrutiny.³²

Similarly, the president or governor will also have strong long-term interests at stake in nominating persons of stature. Appointing governors or presidents will be interested in seating well-reputed judges to establish their own legacy as someone who took nominations powers seriously and worked to favorably influence the administration of justice. The president or governor may also want to appoint a jurist who has the vision and the ability to influence the law for future generations.

30. *Confirmation Hearings on Federal Appointments: Hearings Before the S. Comm. on the Judiciary*, 107th Cong. 707-35 (2002).

31. Attitudinalists might suggest that appointing someone sharing a similar legal vision is simply an attempt to achieve favorable political results through judicial decisions. But the equation is more complicated. A conservative governor, for example, is likely to choose to appoint a conservative judge over a liberal candidate because they think the conservative judge is right on the law. Conservative governors, then, choose judges based on what they believe to be merit; the potential to obtain favorable political results through judicial decisionmaking is not, in effect, anything more than a welcome by-product.

32. This is not to say that a president or governor will never appoint a judge because of immediate political concerns. President George W. Bush, for example, may have nominated Harriet Miers to the United States Supreme Court because he was particularly interested in having a judge who strongly supported his positions on expansive presidential power.

The ranking system set forth by Professors Choi, Gulati, and Posner neither looks at case results nor reputation in defining what constitutes a good court.³³ They dismiss the former as inserting the political component that an objective evaluation is designed to avoid³⁴ and they dismiss the latter because determining judicial reputation would inevitably require surveying “experts” such as law professors or bar organizations who may use surveys to advance their own political agendas.³⁵ Rather, the authors’ rankings look solely to who scores well on the chosen proxies for productivity, influence, and independence. And although the authors, to be sure, are aware that reliance on these factors alone may omit other important aspects of judicial behavior,³⁶ their ranking system, in the end, weighs only these three proxies.

We agree with the authors’ conclusion that neither interest groups nor expert rankings provide objective assessments of judicial performance. We are concerned, however, that the factors measured in *Judicial Evaluations* are too narrow to be granted the presumptive weight the authors suggest. Given the potential real-world impact of this ranking system, it becomes necessary to investigate whether the rankings in *Judicial Evaluations* actually work to reflect (and promote) judicial quality. We turn to this task in the next Section.

III. THE CHOI, GULATI, AND POSNER MEASURES

It is not our purpose to raise every conceivable objection to the methodology employed by the authors in *Judicial Evaluations*. Professors Choi, Gulati, and Posner are refreshingly candid in

33. Choi et al., *supra* note 1, at 1320–26.

34. *Id.* at 1315–16 (“Because of the difficulty of identifying the principal’s preferences, we cannot very easily evaluate judges on the basis of case outcomes.”).

35. Many scholars have critiqued the ABA’s ratings of judicial nominees. *See, e.g.*, Bruce Fein, *Let’s Lower the ABA from Its Pedestal*, LEGAL TIMES, Sept. 14, 1992, at 28; *see also* James Lindgren, *Examining the American Bar Association’s Ratings of Nominees to the U.S. Courts of Appeals for Political Bias, 1989–2000*, 17 J.L. & POL. 1, 2 (2001) (“In recent years, the ABA’s role in rating judges has become increasingly controversial . . .”); Laura E. Little, *The ABA’s Role in Prescreening Federal Judicial Candidates: Are We Ready to Give up on the Lawyers?*, 10 WM. & MARY BILL RTS. J. 37, 37–39 (2001) (“The ABA’s explicitly controversial positions have surely contributed to its public relations problems and have magnified suspicions that the ABA uses judicial evaluations to implement policy objectives under the whitewash of ‘judicial fitness.’”); William G. Ross, *The Supreme Court Appointment Process: A Search for a Synthesis*, 57 ALB. L. REV. 993, 1026 (1994) (“The perception . . . has inspired criticisms of its prominent role in the appointment process.”).

36. Choi et al., *supra* note 1, at 1320–26.

recognizing potential methodological objections to their study and they have been remarkably creative in striving to control for state-specific factors in compiling their results. Rather, our intent is to critique *Judicial Evaluations* with a broader brush. Specifically we ask three deliberately broad questions: (1) do the measures of productivity, influence, and independence accurately assess judicial quality; (2) does the data collected in the study create meaningful proxies for these qualities; and (3) does ranking based on performance on these measures encourage beneficial judicial behavior?

A. Do Productivity, Influence, and Independence Reflect Judicial Quality?

Although it obviously depends on how they are defined and measured, we assume that productivity, influence, and independence are positive judicial qualities. We suggest, however, that assessing judicial quality by measuring these factors may not be particularly fruitful and, perhaps, may be counterproductive.

Consider, for example, what types of judicial decisionmaking are not captured by the study's measures. As the authors acknowledge, their study does not measure whether courts or judges decide cases on the basis of policy preferences, although the authors themselves suggest that deciding cases in that way is improper.³⁷ Similarly, the study does not pick up judicial corruption or malfeasance.³⁸ Nor does their study, in examining productivity, influence, and independence alone, uncover the judicial behavior discussed in Professors Thomas Brennan, Lee Epstein, and Nancy Staudt's contribution to this Symposium. That study suggests that judges at times act like voters and punish or reward the government based on economic conditions that may be extraneous to the legal issues before them.³⁹ Therefore, we wonder whether Professors Choi, Gulati, and Posner's study can grant a presumptive marker of excellence or high quality if it allows a court to maintain a high ranking even when its members decide cases or behave in ways that the authors and most observers acknowledge

37. *Id.* at 1325.

38. *Id.* (discussing the fact that the study does not pick up instances of bribery).

39. See Thomas Brennan, Lee Epstein & Nancy Staudt, *Economic Trends and Judicial Outcomes: A Macrotheory of the Court*, 58 DUKE L.J. 1191, 1194 (2009).

are inappropriate.⁴⁰ Equally significant, because of data limitations, their study ignores aspects of judicial behavior that are arguably more important than the ones proxied, such as integrity, fairness, open-mindedness, thoroughness, and temperament. And it neglects the contributions to the law that judges may make outside opinion writing, including judicial administration and lower court supervision, serving on government commissions, teaching at law schools, writing scholarly articles, and training other judges. The data can only provide so much information.

Professors Choi, Gulati, and Posner, of course, recognize this criticism and readily acknowledge that their measures “capture some, but not all, aspects of judicial quality.”⁴¹ But they do not concede that it is therefore misleading to focus rankings solely on their performance measures. As noted, they argue that their rankings, although not definitively capturing judicial quality, at least create a presumption that courts that do well on their scores are superior to those that do not. They state, “It would be a mistake to believe that small differences in measured outcomes reflect significant differences in quality. But where the differences are large, it is likely that the lower-ranked judges or courts are inferior, unless a good reason exists to explain the difference.”⁴² We are not convinced. If productivity, influence, and independence represent only a narrow range on the list of attributes that constitute a good judge or court, then creating a presumption that a court that scores well on these factors is a high quality court may be a classic example of the tail wagging the dog.

The authors strive to avoid this criticism by, in effect, understating the impact of their work. They contend that because their study only seeks to establish a presumption of judicial quality rather than a full measure, anyone who disagrees with their results can simply come forward and explain why a court that fares relatively poorly on their measures should otherwise be deemed a quality court. To the authors, the import of their study is in its effect as an

40. It might be argued that if a court's decisions are too politically based they will not be cited outside of that court's jurisdiction and would therefore fare poorly on the influence measure. That would only be true, however, if other courts were also not being swayed by political concerns.

41. Choi et al., *supra* note 1, at 1319.

42. *Id.* at 1319–20.

“information-forcing” device.⁴³ The political power inherent in rankings, however, is not that nuanced. A claim by a court’s defender that the results of a purportedly objective study should be discounted because that court has chosen to devote its energies toward activities that are outside the objective measures or that the court did not do well in relation to other courts because of the idiosyncratic nature of its caseload would have little political resonance and would simply sound defensive. Similarly, because the public has little or no basis for distinguishing between true and false claims about nonquantifiable factors as reputation, stature, fairness, and judicial temperament, any claim by a court’s defender arguing that the court’s poor performance on the study’s rankings is outweighed by its excellence in these other areas would sound hollow.

The compelling information in this political dialogue and the information that, in Professors Choi and Gulati’s words, has the “ease of transmission and understandability,”⁴⁴ is the ranking number and not the long excusatory explanation that follows. The study’s objective rankings, in short, create a hard presumption that would be extraordinarily difficult to rebut.

To all of this, one other important factor should be added. The claim that the chosen proxies for productivity, influence, and independence should presumptively trump other judicial qualities requires, at the least, that those proxies are precisely measured. But whether *Judicial Evaluations* accurately gleans productivity, influence, and independence is a matter to which we now turn.

B. Does the Data Collected in Judicial Evaluations Accurately Capture and Compare Productivity, Influence, and Independence?

1. *Productivity.* Professors Choi, Gulati, and Posner purport to capture productivity by counting the number of opinions issued by each state supreme court. But whether the number of published opinions should be equated to productivity depends in large part on what goals the productivity measure is designed to serve.⁴⁵ The

43. For a general discussion of the value of information-forcing devices, see Scott Baker, Mitu Gulati & Stephen Choi, *The Rat Race as an Information-Forcing Device*, 81 IND. L.J. 53, 78 (2006).

44. Choi & Gulati, *Choosing the Next*, *supra* note 2, at 39 n.29.

45. See, e.g., Daniel A. Farber, *Supreme Court Selection and Measures of Past Judicial Performance*, 32 FL. ST. U. L. REV. 1175, 1177–78 (2005) (suggesting that the productivity measure does not clearly measure beneficial judicial behavior).

authors argue that productivity is an important measure of judicial quality because it reflects whether the courts are fulfilling their obligation to resolve cases.⁴⁶ But, in fact, as the authors acknowledge, the relationship between publishing opinions and resolving cases is not immediately apparent. Courts also resolve cases by issuing unpublished opinions. Yet the authors do not credit these forms of case disposition in their opinion counting, claiming that because unpublished opinions are written by secondary personnel they do not reflect the work of the court itself.⁴⁷ Why not? If what is being rewarded by the productivity measure is the court's actions in deciding cases, it should not matter how those cases are resolved.

The authors also defend the published opinions proxy on the ground that written opinions allow judges to share their reasoning with the parties and other judges thereby bettering the system of justice.⁴⁸ But more opinions do not necessarily lead to a better jurisprudence. Multiple decisions on the same subject can often serve to convolute, rather than clarify, legal rules.⁴⁹ Authoring a single opinion that provides a clear rule of decision may therefore be a more productive use of judicial resources than writing multiple decisions on the same subject.

Further, judges are productive in a variety of ways other than merely increasing the quantity of opinions that they publish. In addition to issuing unpublished opinions, courts may also devote considerable time and resources to reading lengthy trial records, to reforming judicial rules and practices, to supervising the state's lower courts, to writing scholarly articles, to training new judges, to facilitating settlements, and to serving on government commissions. None of these efforts are captured by the study's productivity proxy.

46. Choi et al., *supra* note 1, at 1320–21.

47. *Id.* Presumably, the authors do not value productivity merely as a surrogate for hard work. Hard work may be an admirable trait but it should not be relevant to a ranking system unless it leads to results that benefit overall judicial quality. In any case, quantity of production does not always equate to individual effort. A judge, after all, may work just as hard as another but, for a variety of reasons, not be able to produce as much written work.

48. *Id.* at 1321.

49. See, e.g., Kirt Shuldberg, *Digital Influence: Technology and Unpublished Opinions in the Federal Courts of Appeals*, 85 CAL. L. REV. 541, 547–49 (1997) (describing research costs and judicial efficiency as rationales for not publishing all opinions); see also ADMIN. OFFICE OF THE U.S. COURTS, ANNUAL REPORT OF THE DIRECTOR OF THE ADMINISTRATIVE OFFICE OF THE UNITED STATES COURTS 1964, at 11 (1964) (recommending for the first time that opinions be published only in cases “which are of general precedential value”).

Finally, relying solely on the number of published opinions ignores other key (and perhaps competing) proxies, such as timeliness in deciding cases, that also reflect a court's productivity. Consider two courts. In a given year, Court *A* hears 300 cases. It writes opinions in 150 cases, with each opinion taking, on average, 100 days to formulate and decides the other 150 by unpublished decisions, which take, on average, 30 days each to formulate. Court *B*, in turn, hears 300 cases, writes 50 opinions, and decides the other 250 cases by unpublished decision. Which court is more productive? Court *A* writes more opinions; so the authors' study would give that court a higher rating. But Court *B*'s speed of case disposition is much faster.⁵⁰ Given the importance of timeliness to the administration of justice,⁵¹ why should not Court *B* be considered the more productive court?⁵²

2. *Influence.* The influence proxy employed by Professors Choi, Gulati, and Posner is also problematic. Whether citation counts generally measure influence or opinion quality has been the subject of much debate, and we need not review that literature here.⁵³ Even if citation counts do shed some light on influence or quality, however, they are just part of the picture. There is much more to opinion quality than the number of judicial citations an opinion generates. A landmark opinion in a relatively narrow area of law, for example, may be cited less than a mediocre opinion in a frequently litigated substantive area simply because there are fewer opinions generated in that area.⁵⁴ Or a court may write a groundbreaking opinion adopting a minority approach to a legal issue; but that opinion will not yield many citations if other states decide not to adopt the minority approach.

50. On average Court *B* decides each case in 41.6 days; on average Court *A* decides each case in 65 days.

51. We are reminded of the adage "Justice delayed is justice denied," often attributed to William Gladstone and made famous as a plea to courts to be efficient to administer justice properly. See, e.g., John Paul Stevens, *In Memoriam: Judge Donald P. Lay*, 92 IOWA L. REV. 1551, 1585 (2007) (praising Judge Lay's speed and effectiveness in resolving cases).

52. Indeed, one of the problems inherent in the authors' use of published opinions as a proxy for productivity is that it discourages courts from issuing decisions quickly because writing opinions takes more time.

53. See, e.g., Frank B. Cross & Stefanie Lindquist, *Judging the Judges*, 58 DUKE L.J. 1383, 1391–93 (2009); William M. Landes, Lawrence Lessig & Michael E. Solimine, *Judicial Influence: A Citation Analysis of Federal Courts of Appeals Judges*, 27 J. LEGAL STUD. 271, 272–75 (1998).

54. See Cross & Lindquist, *supra* note 53, at 1392 ("[S]ome types of cases are simply more common and therefore more likely to receive citations than other types of cases.").

Further, there are particularly strong reasons to question the use of citation count for its accuracy in comparing influence or quality across state systems. To begin with, we suspect that citations to out-of-state courts are likely to come from the lawyers in the case, not from in-state judges persuaded by the power of the reasoning of their out-of-state colleagues. And, needless to say, the reasons that lawyers cite out-of-state cases are also not primarily related to the strength of the court's reasoning. They choose to cite cases because they welcome virtually any authority that supports their legal position.

Finally, high citation counts may be attributable to developments in law unrelated to the strength of the judiciary. Citations to out-of-state courts frequently occur when there is no case law on point in the home jurisdiction, that is, when the legal issue in dispute is a matter of first impression. This inevitably favors courts with innovative legislatures, litigants, or state executives because those entities are likely to trigger first-impression litigation and courts that rule first in an area are likely to be more frequently cited, whether or not the opinion itself is well reasoned.

3. *Independence.* A critical question regarding any independence measure is "Independence from what?" The authors' study examines a judge's independence by counting the number of times a judge writes an opinion in opposition to another judge of his own party. As one of us has previously argued, independence defined in this way may actually measure a lack of quality in judging, including deficiencies in collegiality and leadership, and a propensity for judicial activism outside the ideological mainstream.⁵⁵ Further, dissenting opinions can also be harmful to the court's general administration of justice as they require a court to spend more time and devote more resources to the case in question at the expense of other cases.

Even more critically, the authors' exclusive focus on intracourt voting behavior in assessing independence ignores the nature of state judicial selection processes. Unlike the federal court system in which the judges enjoy life tenure, state court judges are generally subject to reelection or reappointment and therefore need to have the political support of governors, political parties, and contributors. The real tests of a judge's independence in the state courts, then, is more likely how

55. William P. Marshall, *Be Careful What You Wish for: The Problems with Empirical Rankings to Select Supreme Court Justices*, 78 S. CAL. L. REV. 119, 127-28 (2004).

the judge responds to the pressures from these interests.⁵⁶ After all, it is governors, political parties, and contributors who, unlike fellow judges, have the ability to affect a judge's tenure and retention on the bench. The authors' study, however, does not measure how the judge votes in relation to positions of these entities. As such, the authors' proxy of counting the number of opinions a judge writes in opposition to judges of their own party may not provide much insight into judicial independence at the state court level.⁵⁷

C. Do the Authors' Measures of Productivity, Influence, and Independence Encourage Salutory Judicial Behavior?

As noted previously, ranking systems inevitably encourage their subjects to engage in behavior that increases their scores. Thus, *Judicial Evaluations* needs to be examined in this context as well—do the measures used in the study encourage beneficial judicial behavior? That is, would the quality of state courts be improved if the courts issued more published opinions, worked to have their opinions more widely cited out of state, and had judges more frequently write opinions in opposition to others in their political party?

Although we have no doubt that some judicial observers might believe that state court jurisprudence would be improved if judges responded to these incentives, we are skeptical that many would find that the major problems in state court jurisprudence are too few opinions, too few out-of-state court citations, or too few dissents. Rather, we suspect that observers would be more concerned that judges too readily insert their policy preferences into their decisions⁵⁸ or are too beholden to political and special interests.⁵⁹ But as we have discussed, nothing in the authors' study deters judges from improperly inserting their policy preferences into their judicial decisions or being unduly influenced by the policy agendas of their

56. See generally Joanna M. Shepherd, *Money, Politics, and Impartial Justice*, 58 DUKE L.J. 623 (2009) (demonstrating statistically that the views of donors and those who appoint judges affect judges' voting decisions).

57. On this point, it is notable that the authors' independence findings do not correlate with the Chamber surveys that ask lawyers to rate courts for judicial impartiality. See Choi et al., *supra* note 1, at 1355–59.

58. See, e.g., Lawrence B. Solum, *Judicial Selection Versus Character*, 26 CARDOZO L. REV. 659, 661 (2004); Lawrence B. Solum, *The Virtues and Vices of a Judge: An Aristotelian Guide to Judicial Selection*, 61 S. CAL. L. REV. 1735, 1741 (1988).

59. See, e.g., Stephen J. Ware, *Money, Politics, and Judicial Decisions: A Case Study of Arbitration Law in Alabama*, 30 CAP. U. L. REV. 583, 591–95 (2002) (explaining that interest groups may buy judge-made policy at the state level with campaign contributions).

political patrons, financial supporters, or reappointing agents.⁶⁰ Nor do the incentives created by the authors' measurements respond to other frequent criticisms of the courts, including inefficient judicial administration and long delays in deciding cases.⁶¹ Indeed, the study places a premium on writing more published opinions and dissents, which only exacerbates these problems. Thus, there is a significant question as to whether the ranking system in *Judicial Evaluations* also misses the mark by not creating a system of incentives that better captures the elements of what should be encouraged and discouraged in judicial behavior.

IV. CRAFTING A WIDER JUDICIAL RANKING LENS?

The fact that some problems can be identified in the *Judicial Evaluations* methodology does not take any force away from its authors' observation that judicial rankings are here to stay. Nor does it weaken their case that developing objective measures of judicial quality would be beneficial in providing counterweights to competing agenda-driven systems. It does suggest, however, that an alternative lens for evaluating judicial performance needs to be developed.

The problem is that there are endemic weaknesses in any ranking system. Objective measures such as the one employed by Professors Choi, Gulati, and Posner inevitably miss essential nonquantifiable aspects of judicial quality. The data are just not sufficiently rich.

Expert evaluations are beset by their inherent subjectivity and susceptibility to ideological bias. Agenda-driven surveys such as that used by the Chamber do not, by definition, produce politically neutral results.

We believe an ideal ranking system would serve two important functions. First, it would provide necessary counterweight to agenda-driven rankings, ideologically tinged expert evaluations, and efforts to hide politically driven nominations behind claims of judicial excellence. Second, an ideal ranking system would provide incentives and disincentives for judicial behavior and accordingly would provide a potentially valuable tool for improving judicial performance. A

60. See Shepherd, *supra* note 56, at 629 (“[U]nder some retention methods, judges’ voting is associated with the political preferences of those who will decide whether the judges keep their jobs.”).

61. On the value of speedy courts, see *supra* note 51.

ranking system should, therefore, be adopted with these goals in mind.

In this respect, we propose, perhaps counterintuitively, that a better way to further these goals would be to first encourage multiple interest groups to rank judges and then to examine who does well across a wide range of subjective rankings. Although any particular subjective evaluation may be unreliable, it may be possible to discern important attributes by comparing different subjective evaluations and taking into account the sources of those evaluations. If a court has performed well in the Chamber study standing alone, for example, this may simply reflect that commercial lawyers perceive that court to be probusiness. But if that court also scores well with organized labor, environmental organizations, and trial attorneys, the rankings take on a different light. They suggest instead that the court's reputation derives from aspects of judicial performance that transcend politics and ideology, perhaps including qualities such as fairness, character, integrity, intellect, and efficiency.

At the same time, if a court ranks well with only the groups on one side of the political spectrum, it sends precisely the opposite message. A court that scores well with only one type of interest group while earning failing marks from others is subject to the immediate criticism that the court is only a shill for the interest group that has evaluated it favorably. In this way, the fact that a court scores well with a particular interest group can effectively be turned into a political liability rather than an advantage. Multi-interest group rankings can therefore provide effective counterspin to single interest group rankings and provide counterweight to agenda-driven quality claims.

Multi-interest group rankings also serve the goal of encouraging beneficial judicial behavior because a judge, to improve their overall rankings, will need to be favorably perceived across a broad spectrum of interests. This motivation should encourage the judge to be, among other things, fair, thoughtful, open-minded, tempered, and solicitous of opposing views—to be, in short, a better judge.

To be sure, utilizing multi-interest group rankings is no panacea. It is not clear, for example, how the rankings of the varying interest groups should be solicited or cumulated, if at all. Nor is it immediately apparent how the groups whose rankings are to be considered in an overall appraisal should be identified. And any cumulative ranking system will also raise inevitable questions of balance. Does a cumulative ranking system that incorporates one

survey each from environmental, labor, consumer, and business interests reflect a balanced measure, or is it decidedly antibusiness because three of those four groups would not be expected to be particularly sensitive to business concerns?

Finally, a multi-interest group approach may also invite strategic behavior. A liberal group may very well believe certain conservative jurists are excellent judges, for example, but nevertheless give them low scores to weaken those judges' candidacy for judicial promotion. Indeed, to the extent ratings are used in selecting Supreme Court Justices, as Professors Choi and Gulati suggested in an earlier article, strategic behavior of this type might quickly become the norm.

Nevertheless, we believe that expanding the field of avowedly subjective judicial rankings may be a more effective mechanism for evaluating judicial quality and for promoting beneficial judicial behavior than a measure that too narrowly focuses on uncontested objective factors. And given that judicial rankings, for better or worse, are unlikely to disappear, multigroup rankings may be the only effective way to counter the political agendas of interest groups acting on their own.

CONCLUSION

In their project, Professors Choi, Gulati, and Posner tackle a difficult task: ranking the state supreme courts on objective measures for which data is readily available. They employ a transparent and easy to understand methodology, attempting to measure judicial productivity by counting a court's number of published opinions, judicial influence by counting out-of-state court citations, and judicial independence by counting the number of opinions a judge writes in opposition to another judge of her own political party. Recognizing that productivity, influence, and independence alone do not fully capture judicial quality, they suggest that performance on these measures should set a presumption of judicial excellence that can be rebutted by the inclusion of additional information to explain the results. The rankings, as they see it, serve both to provide an initial insight into judicial quality and an information-forcing device for more complete appraisals.

Our response to the project is not the obvious one: Quality cannot be measured so do not even try to grade the state courts. Rather, recognizing, as the authors do, that, like it or not, judicial rankings are here to stay, we pose a series of questions.

Do most observers really care about productivity, influence, and independence when it comes to state supreme courts? Or do they care more about traits in judges that cannot be objectively measured such as fairness and judicial temperament?

Even if observers do care about productivity, influence, or independence, does scoring well on the proxies used by the authors in measuring these qualities accurately measure excellence in the attributes themselves? After all, the authors use only one proxy for each attribute considered. Is the number of published opinions alone, for example, an effective proxy for judicial productivity? Or would using more proxies provide a better measure?

Can the presumption of quality triggered by the ranking effectively be rebutted by counterclaims about the judicial qualities undetected by the Choi, Gulati, and Posner framework? It may not be enough that rankings will force information. They probably will. The question is whether the information presented in response to the rankings will actually be able to rebut the presumptions they create. We do not believe they will.

Finally, do the measures encourage judges to act in a way that enhances judicial quality? If not, as we suspect, then it may be that the rankings will effectively undermine rather than improve judicial performance.

In the end, perhaps it is best to admit that direct objective evaluation is difficult here and that the better alternative is to encourage more unabashedly agenda-driven rankings by interest groups like the Chamber. First, because their motivations will generally be clear, any weight given to these subjective rankings can be discounted accordingly (as long as competing rankings are available). Second, if a judge or court performs well across a wide range of interest groups, it may provide meaningful and useful insight about judicial quality that transcends the value of any appraisal offered by one study alone.