

2005

Supreme Court Selection and Measures of Past Judicial Performance

Daniel A. Farber
daf@daf.com

Follow this and additional works at: <http://ir.law.fsu.edu/lr>



Part of the [Law Commons](#)

Recommended Citation

Daniel A. Farber, *Supreme Court Selection and Measures of Past Judicial Performance*, 32 Fla. St. U. L. Rev. (2005) .
<http://ir.law.fsu.edu/lr/vol32/iss4/7>

This Article is brought to you for free and open access by Scholarship Repository. It has been accepted for inclusion in Florida State University Law Review by an authorized administrator of Scholarship Repository. For more information, please contact bkaplan@law.fsu.edu.

FLORIDA STATE UNIVERSITY LAW REVIEW



SUPREME COURT SELECTION AND MEASURES OF PAST JUDICIAL PERFORMANCE

Daniel A. Farber

VOLUME 32

SUMMER 2005

NUMBER 4

Recommended citation: Daniel A. Farber, *Supreme Court Selection and Measures of Past Judicial Performance*, 32 FLA. ST. U. L. REV. 1175 (2005).

SUPREME COURT SELECTION AND MEASURES OF
PAST JUDICIAL PERFORMANCE

DANIEL A. FARBER*

I. INTRODUCTION.....	1175
II. INDICATORS OF JUDICIAL QUALITY	1176
A. <i>Productivity</i>	1177
B. <i>Citations as a Measure of Opinion Quality</i>	1178
C. <i>Independence</i>	1179
III. PROBING THE MEASURES OF OPINION QUALITY.....	1180
A. <i>Power Laws and Normal Distributions</i>	1181
B. <i>Invocations</i>	1184
C. <i>Citations</i>	1188
IV. THE RELEVANCE OF OBJECTIVE MEASURES OF PERFORMANCE	1192
V. CONCLUSION	1195

I. INTRODUCTION

*Can we measure the quality of a judge’s past performance? Should the President weigh this factor in making Supreme Court appointments?*¹ I will argue that the answer to both questions is the same: “Yes, but with significant qualifications.” Yes, we do have some indices for assessing judicial performance—but those measures are rough and incomplete. Yes, the President should consider past judicial performance and its objective indicators—but only to a limited degree.

These comments are prompted by two important recent articles by Stephen Choi and Mitu Gulati.² They make a more ambitious claim: that we can measure judicial quality well enough to make it the predominant element in Supreme Court appointments. In their initial article, they proposed a tournament, an objective contest among federal appellate judges in which the prize is a Supreme Court appointment.³ In their follow-up article, they run an initial tournament using recent data about federal appellate decisions.⁴ They hedge their conclusions about “who is the fairest in the land,” but the clear

* Sho Sato Professor of Law, University of California at Berkeley.

1. In my view, the Senate is probably entitled to consider the same factors as the President in determining whether the nomination is suitable. But the role of the Senate in judicial confirmations is controversial and beyond the scope of this Essay.

2. Stephen Choi & Mitu Gulati, *A Tournament of Judges?*, 92 CAL. L. REV. 299 (2004) [hereinafter Choi & Gulati, *Tournament*]; 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, *Empirical Ranking*].

3. Choi & Gulati, *Tournament*, *supra* note 2.

4. Choi & Gulati, *Empirical Ranking*, *supra* note 2.

import seems to be that Richard Posner should be the next Supreme Court appointment.⁵

While the tournament idea is intriguing, it is far too ambitious. In my view, a conscientious President probably would not choose Justices on the basis of such a tournament, partly because we cannot measure professional merit well enough and partly because professional merit is not the only factor that the President should find relevant. Nevertheless, Choi and Gulati have made a valuable contribution by showing how objective measures of quality can improve the appointments process. Simply making the tournament data available on a regular basis through a reliable information source could help temper partisan claims and provide guidance for the public. Moreover, the data in their follow-up article raises tantalizing questions about the appellate process and, in particular, about the dynamics of judicial influence. Those questions would be well worth pursuing even if the results had nothing to tell us about how to appoint Justices.

II. INDICATORS OF JUDICIAL QUALITY

Supreme Court Justices engage in four major tasks: (1) they vote on which cases to accept, (2) they vote on the outcomes of cases, (3) they write majority opinions, and (4) they write dissents and concurrences.⁶ By contrast, federal appellate judges only engage in two of these tasks.⁷ Federal appellate judges rarely write separate opinions, so the fourth factor is less significant in assessing their performance. Also, nearly all appeals at the circuit level are mandatory, which makes the first factor less significant. This leaves two major tasks for federal circuit judges: voting on the merits and writing opinions. How well a judge performs these tasks on a lower court presumably has some predictive value about potential performance on the Supreme Court.⁸ Choi and Gulati attempt to measure circuit judges' performance of these tasks along three dimensions: productivity, quality of opinion writing, and independence.⁹ Of these, as we will see, the at-

5. *Id.* at 74, 113 tbl.H. I know a number of academics who agree—including some who are liberals or skeptics about Posner's specialty, law and economics—but I do not know of anyone who thinks the prospect is likely.

6. Justices also engage in other tasks, such as considering stay orders or approving changes in federal procedural rules, but these seem less central to the role.

7. The current norm is for Presidents to nominate circuit judges to the Supreme Court. Lee Epstein et al., *The Norm of Prior Judicial Experience and Its Consequences for Career Diversity on the U.S. Supreme Court*, 91 CAL. L. REV. 903 (2003); see also Choi & Gulati, *Empirical Ranking*, *supra* note 2, at 40. Whether this norm is desirable is outside the scope of this Essay.

8. How much predictive value is another question, which we will consider in Part IV, *infra*.

9. See Choi & Gulati, *Empirical Ranking*, *supra* note 2, at 33, 42.

tempt to measure opinion quality via citation rates is probably the most successful.

A. *Productivity*

One of the stated purposes of the tournament is to encourage judges to work harder as they compete for higher rankings and Supreme Court nominations.¹⁰ One of the indices of judicial performance, consequently, is productivity. Since circuit judges have discretion over which of their opinions are published, Choi and Gulati argue that judges with a higher number of published opinions are investing more time and effort in polishing opinions for publication. This is quite plausible, but there are alternative explanations for high publication rates that make these rates a weak indicator of effort.

One possible explanation for a high publication rate could be that a judge has trouble identifying which cases are important enough to serve as precedents and therefore deserve published opinions. Such an inability would be a handicap in a Supreme Court Justice. One of the Justice's tasks is to decide which cases deserve Supreme Court review; thus a judge who cannot accurately gauge the significance of a case is at a disadvantage.

Another possible explanation is self-centeredness. Some judges may consider all of their output to be worthy of preservation for the ages, simply because it is theirs. A related explanation is that a judge may fail to appreciate the costs that publication imposes on others. The two other members of the panel must invest additional time in reviewing and commenting on a published opinion, and additional published opinions increase the size of the pool of authority that lawyers must search. Productivity based on self-centeredness does not speak well of judicial quality or recommend promotion to the Supreme Court.

An alternative explanation for high publication rates is consistent with the idea that some judges invest more effort in perfecting their opinions for use as precedents. Judges may vary in how important they consider this function of opinions, and some judges may focus more heavily on their role as generators of precedents at the expense of their dispute settlement role. A published appellate opinion not only provides a precedent, it resolves a specific piece of litigation. It is important for the functioning of the legal system as a whole that appeals be correctly decided, which means not only that the correct rules of law are announced for the future but that the rules are relevant and correctly applied to the specific case. This may take tedious

10. Choi & Gulati, *Tournament*, *supra* note 2, at 304.

investigation of the trial record. A judge who is willing to cut corners on fidelity to the record can publish more opinions, and those opinions may be well regarded as precedent. Yet, that judge also decreases the ability of the legal system to respond accurately to the facts of cases and degrades the overall performance of the system (besides being unfair to individual litigants). Furthermore, individual cases themselves sometimes have substantial social importance (think of the Microsoft antitrust litigation or of redistricting challenges), so it is important for judges to invest time in scrutinizing the record even if doing so does not increase the opinion's precedential value.

A final, and even simpler, explanation is that some judges need not invest a high level of effort in order to get a high publication count, either because they themselves are facile writers or because they are able to recruit and motivate highly productive law clerks. Being able to write easily, or successfully motivate clerks to do so, is a valuable trait. It is not, however, a gauge of effort. Moreover, judges are probably either good writers when they are appointed (or good supervisors) or they are not; either way, providing incentives is unlikely to make much difference.

The bottom line is that productivity may measure good judicial traits such as effort or writing ability or bad judicial traits such as self-centeredness or sloppy treatment of facts. Without knowing more, we cannot be sure of what we are measuring or whether it is something we want to encourage.

B. Citations as a Measure of Opinion Quality

Are citation counts, of any kind, a gauge of quality? In the context of appellate judges, it seems plausible to think so. For citations to be unrelated to quality, we would have to assume that judges are just as likely to cite and rely on muddled or ill-reasoned opinions as on cogent and logical ones. This would indicate a serious problem with the functioning of the federal judiciary. Presumably, judges tend to cite cases which best support their arguments or most require distinguishing or rebuttal. Consequently, they should favor strong opinions over weak ones in their citation practice.

Yet we should not expect citations to be a very precise measure of quality. Too many other factors may influence how often a judge's opinions are cited: the "luck of the draw" (how many significant cases are drawn by panels on which that judge sits), the author's reputation (which may imperfectly reflect actual quality), the quality of the judge's law clerks in a particular year, ideological or personal affiliations (or antipathies) between judges, and just plain luck (perhaps an opinion gets cited more often because its name is short and memora-

ble or because the facts are striking).¹¹ For these and other reasons, it seems right to say that “[c]itations are at best a crude and rough proxy for measuring influence.”¹² In turn, influence may only imperfectly reflect quality.

Moreover, we may not want circuit judges to strive to maximize their citation rates. In seeking to increase their influence, judges may reach for broader holdings by ignoring the factual nuances of specific cases. Or in trying to write influential opinions, they may simply not pay much attention to the litigation record. They might also be likely to insert unnecessary dicta or to address issues not raised by the lawyers or addressed by the trial judge. None of these seem to be particularly beneficial behaviors.

Although these points suggest the need for caution in relying on citation counts, it remains true that citation counts tell us something important about how useful opinions are to other judges. A judge whose opinions are consistently useful to others is probably doing something right, while a judge whose opinions are rarely cited is probably performing badly.

C. Independence

“Knee-jerk” is not usually considered a good adjective as applied to judges. We presumably do not want judges who always vote for the most liberal or conservative result without considering specific legal arguments. It is probably even worse to have judges who always vote with members of the same political party without regard to the merits of cases. Thus, Choi and Gulati are right to include independence as one of their measures, but their measurement of this characteristic is problematic.

Choi and Gulati measure independence by counting disagreements with other judges appointed by Presidents of the same political party.¹³ This does not seem to be a particularly apt measure. A Republican-appointed judge who often votes on the opposite side from others may be more moderate, but equally well could be an ideological extremist who is parting company with less ideological colleagues. Moreover, since we are talking about relatively small num-

11. Also, the citation figures do not seem to measure quality for members of the D.C. Circuit, who tend to get fewer citations because their workload is specialized, centering on administrative law. See Choi & Gulati, *Empirical Ranking*, *supra* note 2, at 40 n.30 (recognizing that the different nature of the D.C. Circuit's docket from that of the other circuits might make total citations a less meaningful comparison, but including D.C. Circuit judges in the tournament because other citation figures, such as citations to top twenty opinions and citations by law review, might be more meaningful).

12. William M. Landes et al., *Judicial Influence: A Citation Analysis of Federal Courts of Appeals Judges*, 27 J. LEGAL STUD. 271, 271 (1998).

13. See Choi & Gulati, *Empirical Ranking*, *supra* note 2, at 63.

bers of judges in each circuit, it is hard to control for other variables. Some judges may have been appointed by members of the same party but have sharp personality clashes with each other, or one may be markedly more competent than the other. This could lead to frequent disagreements but would not tell us much about whether a judge is voting on a purely ideological basis.

More fundamentally, independence is not a neutral concept. What one of us sees as an ideologue, another may see as a person of principle. Judges who vote on the basis of expediency rather than principle may disagree frequently with their more principled colleagues. Or to put it another way, one might view “independence” as a euphemism for “inconsistency” and hence as an undesirable trait.

Deciding how much weight to give to independence or how to measure it requires a normative judgment. Formalists and pragmatists may have different concepts of what kind of independence is required by judges. For the formalist, it is the ability to strictly follow a set of rules, regardless of pressure to reach a more congenial result. For the pragmatist, it is the ability to be open to opposing arguments, consider the implications of a decision, and reach a balanced conclusion. These translate into very different conceptions of judicial independence.

III. PROBING THE MEASURES OF OPINION QUALITY

Of the three basic criteria discussed in Part II—productivity, opinion quality, and independence—it is the second factor that seems most useful in assessing judicial performance. As we have seen, productivity is an ambiguous quality, which might reflect either well or badly on judicial performance. Independence is normatively charged and hard to measure. Thus, the remainder of this Essay will focus primarily on opinion quality as the measure of judicial performance. Obviously, other highly desirable traits may not be captured by this measure.

Opinion quality is assessed through citation counts.¹⁴ Although citation counts do seem related to the quality of judicial opinions, various different citation counts can be used and the data can be handled in different ways. We need to know more about the processes that generate this data before we can decide on the proper method of measurement. Choi and Gulati have assembled some extremely interesting data, which provide a basis for a more detailed analysis.

This analysis is significant for reasons that go beyond the tournament idea. Citation practices—which opinions are cited, by whom, and for what purpose—are a basic part of the legal system. But we

14. *See id.* at 48-49.

know little about the processes involved. The data presented by Choi and Gulati help shed some light on these broader questions.

A. *Power Laws and Normal Distributions*

In trying to understand the process that produces citation rates, one of the most important clues is the shape of the statistical distribution. Different kinds of processes characteristically produce different sorts of distributions. For purposes of this Essay, we are primarily interested in two types: the normal distribution and the power-law distribution.

The best-known statistical distribution is the “bell curve,” or normal distribution, which is often associated with the idea of random variation.¹⁵ Bell curves are associated with a certain type of random process. According to a basic statistical law, “the sum of a large number of independent random variables will be approximately normally distributed almost regardless of their individual distributions.”¹⁶ Thus, when many small factors combine to produce a result and each factor has an element of randomness, the result is likely to be a normal distribution.

In assessing how close a distribution comes to being normal, there are three useful parameters. The first involves the center of the distribution. For the normal curve, the median, the mean (what most people mean by the “average”), and the mode are identical. The second parameter is skew, which measures asymmetry. A normal curve is symmetrical rather than being stretched in one direction. The skew parameter is zero for the normal distribution. A high skew means that a distribution is bunched on one side and stretched out on the other. The third parameter is called kurtosis. Kurtosis measures whether a curve is flattened out or unusually peaked, compared with the normal distribution. The normal distribution has a kurtosis of three.¹⁷

Apart from the normal curve, we are also interested in another type of distribution—the power-law distribution. Complex, nonlinear systems have a characteristic distribution of outcomes: a “high frequenc[y] of small fluctuations, punctuated by the occasional large shift[] in system conditions.”¹⁸ Rather than following the familiar normal distribution, outcomes in complex systems often follow what

15. See generally M.G. BULMER, *PRINCIPLES OF STATISTICS* 108-15 (Dover ed. 1979) (1965).

16. *Id.* at 109.

17. See *id.* at 63-65.

18. J.B. Ruhl, *Thinking of Environmental Law as a Complex Adaptive System: How to Clean Up the Environment by Making a Mess of Environmental Law*, 34 *HOUS. L. REV.* 933, 952 (1997).

are called power laws¹⁹—that is, the frequency of an event is often given by its magnitude taken to a fixed negative exponent.²⁰

Some additional explanation of power laws may be helpful for the non-mathematically inclined. Albert-László Barabási, a physicist who studies complex networks, explains how power laws work. Contrasting power laws with the normal curve governing characteristics such as human heights, he points out that a frequency distribution “following a power law is a continuously decreasing curve, implying that many small events coexist with a few large events.”²¹ For example, “[i]f the heights of an imaginary planet’s inhabitants followed a power law distribution, most creatures would be really short. But nobody would be surprised to see occasionally a hundred-foot-tall monster walking down the street.”²² Such “outliers” are much less likely when a normal distribution is involved. Or, in more technical terms, “[b]ell curves have an exponentially decaying tail, which is a much faster decrease than that displayed by a power law.”²³

Caselaw can be considered a network of cases linked by citations. Power laws seem characteristic of complicated networks. Some examples include the World Wide Web, where the number of links to a particular site follows a power law; the number of citations to a given physics paper; and even the number of other actors with whom a given Hollywood star has appeared.²⁴ On the Web, for example, about ninety percent of a sample of two hundred million web pages are the targets of ten or less links, while about *three* pages had roughly a million other pages linking to them.²⁵ Similarly, students of bibliometrics refer to Lotka’s law, under which productivity follows an inverse-square relationship: if one hundred scholars in a group each publish one paper annually, then twenty-five will publish two, eleven will publish three, six will publish four, and so forth, with a single scholar producing ten papers.²⁶ In other words, the number of scholars who produce N articles per year is proportionate to $1/N^2$. As it turns out, Supreme Court opinions also appear to follow a power law in terms of their frequency of citation.²⁷

19. For an introduction to power laws, see MANFRED SCHROEDER, *FRACTALS, CHAOS, POWER LAWS: MINUTES FROM AN INFINITE PARADISE* 33-38, 103-19 (1991).

20. See RICARD SOLÉ & BRIAN GOODWIN, *SIGNS OF LIFE: HOW COMPLEXITY PERVADES BIOLOGY* 201 (2000).

21. ALBERT-LÁSZLÓ BARABÁSI, *LINKED: THE NEW SCIENCE OF NETWORKS* 67 (2002).

22. *Id.*

23. *Id.* at 68 n.1.

24. *Id.* at 67-69.

25. *Id.* at 57-58.

26. See Virgil L.P. Blake, *Citation Studies—The Missing Background*, 12 CARDOZO L. REV. 1961, 1962 n.6 (1991) (book review) (citing Linus Ikpaahindi, *An Overview of Bibliometrics: Its Measurements, Laws and Their Applications*, 35 LIBRI 163, 171 (1985)).

27. Daniel A. Farber, *Earthquakes and Tremors in Statutory Interpretation: An Empirical Study of the Dynamics of Interpretation*, 89 MINN. L. REV. 848, 871-72 (2005).

Several possible mechanisms might result in a power-law distribution for judicial citations. One is the superstar phenomenon.²⁸ Suppose that judges differ in quality. The top judge's opinions are slightly better than everyone else's. So when that judge has written on an issue, her opinion is the one cited by other judges. When she has not written on an issue but the second-best judge has done so, the second-best judge's opinion gets cited instead; and so forth. Small differences in quality can translate into large differences in rankings—just as small differences in GPAs can translate into large differences in class ranks.

We can get some sense of how the superstar phenomenon works with a simple model. Assume that the chances that a particular judge will get to write about a particular issue is $1/a$ (in other words, this is the probability that the judge will sit on the first case in the circuit raising that issue and be assigned to write the opinion). Thus, $1/a$ is the proportion of issues that any particular judge can write precedential opinions about. In this simple model, once an issue has been decided in a given circuit, later opinions cite that as the dispositive opinion in the circuit because it binds future panels; later judges in the circuits have no reason to do any more than cite that case. Then the best judge will write on $1/a$ of the issues, and those opinions will be the exclusive citations when those issues arise, so that judge receives a $1/a$ share of the cites. The second-best judge has a $1/a$ share of the remaining cites, or $(1/a)(1 - 1/a)$. By the time we get to the " $(n + 1)$ judge," the number of remaining cases where no previous judge has written is $(1 - 1/a)^n$, and so that judge gets cited in $(1/a)(1 - 1/a)^n$ of the cases. Note that the ratio between the " n judge" and the " $(n + 1)$ judge" in terms of citations is constant: $(1 - 1/a)$. This is a power-law distribution.²⁹

Note that in this model, the ranking of citations perfectly replicates the ranking of quality, but the scale is distorted, since even a tiny difference in opinion quality means that the top judge will be cited in preference to the other judges in every case where the top judge has written.³⁰ The superstar effect can also appear even if there

28. See Choi & Gulati, *Empirical Ranking*, *supra* note 2, at 72.

29. A fuller model would introduce some complications. Certain judges may sit on circuits where more issues of first impression arise, which would increase their chances of hearing such cases. Similarly, by reason of seniority, some judges may be more likely to get assigned the interesting opinions. Also, if the number of novel issues is not proportionate to caseload, judges from smaller circuits might have an advantage because fewer judges would be splitting up the interesting cases. Conceivably, influential judges might sometimes be cited when they merely follow existing circuit precedent, rather than making new law. Finally, it probably is not true that a higher-ranking judge is *always* cited in preference to a lower-ranking one, so we might need to add some random variation.

30. For a summary of various situations that can produce a superstar distribution, see Mitu Gulati & Veronica Sanchez, *Giants in a World of Pygmies? Testing the Superstar Hypothesis with Judicial Opinions in Casebooks*, 87 IOWA L. REV. 1141, 1181-89 (2002).

are no differences in quality between judges. If consumers tend to cite cases in proportion to their existing popularity, with only a small chance of citing something that has not been cited previously, the distribution of product types will tend to something called the Yule distribution, which is close to being an inverse-square power distribution.³¹

Other processes might produce a normal distribution rather than a power-law distribution. The citation count for a given opinion might depend on a whole series of factors, each of which varies randomly. For example, the list of factors might include:

- (1) The judge's past experience with issues in the field.
- (2) The judge's IQ.
- (3) How much time the judge can devote to the particular opinion, given other responsibilities.
- (4) The quality of the lawyers in the case (it is presumably easier to write a good opinion if the briefs are of high quality and the case has been well litigated).
- (5) Who else is on the panel and how many changes they demand.
- (6) Whether the opinion is well classified by the digest systems, and whether the judge used the right key words that will show up in a computer search.
- (7) Whether the opinion happens to be cited in the next few cases on the issue.

If we assume that each of these and similar factors has an element of randomness, their combined effect is likely to look like a normal distribution. On the other hand, a power-law distribution suggests that some kind of feedback is at work. With these observations in mind, we consider the distributions of various measures of opinion quality.

B. Invocations

We begin with invocations, that is, with the number of times the judge is identified by name in later opinions as the author of a major-

Gulati and Sanchez explain the extreme cases using the "pure talent explanation" (where outcomes simply reflect the enormously greater talent of certain individuals) and the "pure randomness explanation" (where an initial lucky success is "locked in" by a cascade effect). *See id.* at 1183-84.

31. *See* Kee H. Chung & Raymond A.K. Cox, *A Stochastic Model of Superstardom: An Application of the Yule Distribution*, 76 REV. ECON. & STAT. 771 (1994). The Yule distribution can be approximated by $1/[(N)(N + 1)]$, *see id.* at 773, which is bounded above by (and converges with) $1/N^2$. Thus the statement in the text that the Yule distribution is "close to" an inverse-square relationship. In another model, reputation effects and search costs can produce the superstar effect even if there are no quality differences. *See* Moshe Adler, *Stardom and Talent*, 75 AM. ECON. REV. 208 (1985).

ity opinion.³² As we will see, this distribution looks nothing at all like a bell curve. In fact, it is more extreme than one would expect from a power-law distribution.

Table 1 shows that something odd is going on with invocation rates.

TABLE 1
FREQUENCY OF INVOCATIONS³³

TOTAL INVOCATIONS	NUMBER OF JUDGES
0-9	88
10-24	8
25-49	0
50-74	0
75-99	0
100-124	1
125-149	0
150-174	0
175-199	1

(Note that what would otherwise be the first row here (0-24) is split into two, in order to give a better sense of the shape of the distribution.) The distribution consists of two outliers (who are actually Posner and Easterbrook³⁴) along with a large clump. The summary statistics confirm the oddness of this distribution: the kurtosis is a remarkably high 54.7 (recall that the kurtosis of a normal curve is three).³⁵

For purposes of comparison, imagine that the same table showed someone's journal of food expenses. On most days the person lives frugally, spending less than \$10 a day; on other days the person lives more expansively, spending up to \$25. But on two days out of a hundred, the person's food expenses leap over \$150. The obvious inference is that something special happened on those days: either the person splurged at a really fancy restaurant or invited a group of people to dinner. These were obviously not just days when the person

32. For background on invocation rates, see David Klein & Darby Morrisroe, *The Prestige and Influence of Individual Judges on the U.S. Courts of Appeals*, 28 J. LEGAL STUD. 371 (1999).

33. Table 1 was constructed by collating figures from Table F of Choi & Gulati, *Empirical Ranking*, *supra* note 2, at 104-07.

34. Posner had 176 invocations and Easterbrook had 103 during the time span studied by Choi and Gulati. *Id.*

35. *Id.* at 107.

felt a little hungrier and ordered extra food. Similarly, something special seems to be happening with the invocation rates for Posner and Easterbrook.

One possibility is that Posner and Easterbrook are simply judges of an entirely different caliber than their fellows, like two NBA players who have wandered into a lawyers' after-work basketball game. While they may well be exceptionally good judges, however, it seems unlikely that differences in ability level can account for the magnitude of the difference in invocation rates. The third-highest-ranking judge is Guido Calabresi, with twenty-three invocations.³⁶ Posner and Easterbrook are undoubtedly very bright, but are they five times as bright as Calabresi, not to mention the other eminent judges on the list?

Another possibility is the superstar phenomenon—judges invoke only the very best names, given a choice, so a judge who is just a little better than his fellows will get disproportionately more invocations. Without a good model of how reputations develop, it is hard to test this explanation. We can get some sense of its plausibility by assuming that a judge's prestige tracks the number of legal issues on which a judge has written the leading opinion. Under the basic superstar model, this is simply equal to the number of cases on which a judge is cited, since courts are presumed to cite *only* the leading ("best") opinion. But the distribution of invocations does not fit with this explanation. The problem is that no matter how good Easterbrook and Posner are, they first have to sit on a case where an issue is presented in order to write the leading opinions on that issue. There must be a great many legal issues that, for one reason or another, have not been the subjects of opinions by Easterbrook or Posner: either they did not happen to be on a panel on which these issues were raised or someone else wrote the opinion. In these other cases, the superstar phenomenon should lead to large numbers of citations of leading opinions by the third- or fourth-best judges on issues that never reached Posner and Easterbrook. Hence these second-tier judges should have written almost as many leading opinions as Posner and Easterbrook, and their invocation rates should not be drastically lower. But this is not what we see in the data.

Recall that, in the basic model, the superstar effect produces a constant ratio between citations as we move one rank in the ratings. But for invocations, the ratio between judges 1 and 2 is 103/176, or 0.58; the ratio between judges 2 and 3 is 23/103, or 0.22; and the ra-

36. *Id.* at 104 tbl.F.

tio between judges 3 and 4 is $19/23$, or 0.82 .³⁷ Moreover, the first two ratios are hard to account for on the basis of the superstar effect. Consider the 0.58 ratio between the first and second judges. In the basic superstar model,³⁸ the ratio between judges who adjoin in the rank order is $(1 - 1/a)$, and if we set this equal to 0.58 , we get $1/a = 0.42$. (Remember that $1/a$ is the likelihood that a judge will write the controlling circuit opinion on the issue.) But this seems implausibly high—it would require that the top judge happen to sit on the panel (and get assigned the opinion) in 42% of the cases raising issues of first impression in the circuit.³⁹ For similar reasons, the 0.22 ratio between judges 2 and 3 is even harder to square with the superstar model.⁴⁰

Invocations reflect reputations rather than raw citation figures, but it is hard to see why the falloff between reputations should be so steep as we move from the judge who is considered to be the best in the country to the second-best and then to the third-best judge. We cannot exclude this possibility, however, without a better theory of how judicial quality is translated into reputations.

Another plausible possibility is that Easterbrook's and Posner's names are invoked for reasons other than general judicial reputation. Perhaps they are invoked in cases involving their special area of expertise: economic analysis.⁴¹ It would be entirely understandable that these two judges should be the dominant judicial figures on economics issues.

37. The numbers used to calculate these ratios are the total invocations to the top four judges—Posner, Easterbrook, Calabresi, and Wilkinson, respectively—as reported by Choi and Gulati. *Id.*

38. See *supra* text accompanying notes 28-30.

39. Consider a court comprised of N members. What are the odds that a given judge will *not* be on a panel? When we pick the first judge on the panel, the odds are $(N - 1)/N$ of getting someone other than the specified judge. Now there are only $(N - 1)$ judges left in the pool, so the odds of missing the specified judge the second time are $(N - 2)/(N - 1)$. Similarly, the odds of missing the specified judge a third time are $(N - 3)/(N - 2)$. If we multiply these together and simplify, everything cancels out except $(N - 3)/N$, so the chances of missing the specified judge all three times is $(1 - 3/N)$. Hence, the judge will be on $3/N$ of the panels. But Posner is from the Seventh Circuit, and we know that there were ten judges from that circuit in the tournament, see Choi & Gulati, *Empirical Ranking, supra* note 2, at 41 tbl.1, and this does not include post-1998 appointments, senior judges, or judges sitting by designation, all of whom serve on panels. Even for $N = 10$, we would get only a 30% chance of serving on a particular panel, so it is unlikely that Posner would have the opportunity to sit on panels hearing 42% of the cases that raise issues of first impression.

40. With $(1 - 1/a)$ equal to 0.22 , $1/a = 0.78$. Following up the reasoning of the previous footnote, this would require that the judge sit on 78% of the panels hearing issues of first impression. Even in a smaller circuit like the Seventh, any one judge sits on less than half this percentage of cases.

41. Calabresi is also a major figure in the law and economics movement, but his major work in this genre came earlier in his career and was focused on the specific field of torts.

Economics issues are important in some Supreme Court cases, so Posner's and Easterbrook's expertise should count in their favor. But these cases may not be the most important part of the docket, and a President might be warranted in giving this factor relatively low weight. Unreflective use of statistics obscures this question. More generally, this analysis suggests the need for caution in using invocation figures as a measure of quality, because the figures may be skewed by extraneous factors.

As this discussion of invocation statistics shows, it is important to look beyond the raw data in assessing possible measures of judicial quality. While this is most obvious with the outlier cases in the invocation statistics, it is also true of Choi and Gulati's best indicator of opinion quality—outside circuit citations.⁴²

C. Citations

Citations are a more promising gauge of opinion quality. There are, however, several possible measures of citation rates to consider. We begin with Choi and Gulati's "top twenty" citation measure.⁴³ In order to control for the effect of productivity on citation counts, Choi and Gulati identified each judge's top twenty opinions (measured by citations) and then totaled the cites to those opinions. This measure focuses on the impact of the judge's strongest opinions rather than looking at all opinions.

To get a sense of the distributions, it is helpful once again to start with a table.

TABLE 2
CITATIONS TO TOP TWENTY OPINIONS⁴⁴

OUTSIDE CITATIONS TO JUDGE'S TOP TWENTY OPINIONS	NUMBER OF JUDGES
0-100	2
101-200	24
201-300	39
301-400	19
401-500	9
501-600	2
601-700	2
701-800	1

42. Choi & Gulati, *Empirical Ranking*, *supra* note 2, at 49-50.

43. *See id.* at 54, 100-03 tbl.E.

44. Figures collated from *id.* at 100-03 tbl.E.

There are a few things to notice about Table 2. First, unlike the frequency of invocations in Table 1, there are no big gaps in the distribution. While some entries are quite a bit higher than the mean, the numbers climb smoothly to the peak. Second, the distribution is a bit skewed.⁴⁵ Third, it is roughly bell-shaped. These impressions are supported by the formal statistics, which show some skew but a kurtosis of close to three (which would characterize a normal distribution).⁴⁶

Why is this a (roughly) normal curve rather than a power-law distribution? Power laws indicate nonlinearity. For example, we would expect a power law if a judge's top twenty cases were in a feedback loop, where having some of them cited frequently leads to more cites for those cases or for the judge's other top twenty cases. The absence of such a feedback loop (or at least, of any evidence of such a loop) suggests that among the most prominent cases, citations are not reinforcing but rather are based purely on the precedential value of each opinion.

We get a different picture in Table 3, where we look at total citations, rather than just citations to a judge's twenty most-cited opinions:

TABLE 3
TOTAL CITATIONS⁴⁷

TOTAL OUTSIDE CIRCUIT CITATIONS	NUMBER OF JUDGES
101-300	34
301-500	34
501-700	22
701-900	4
901-1100	2
1101-1300	0
1301-1500	2

This does not look remotely like a bell-shaped distribution. Instead, it starts with a large number of judges who get only a few citations and tapers off, at first quickly and then more slowly. This is confirmed by the kurtosis, which is high (5.0)⁴⁸ but not nearly so high as the kurtosis for the invocation statistics (which was about ten times

45. If it were completely symmetrical, some judges would have to have a negative number of citations to balance some of the above-average judges.

46. See Choi & Gulati, *Empirical Ranking*, *supra* note 2, at 103.

47. Figures are collated from *id.* at 94-98 tbl.D.

48. *Id.* at 99.

higher).⁴⁹ This looks a great deal more like a power curve than a normal curve. This should not be surprising—power curves are typical of citation counts in a variety of different fields.

The most notable feature of the power curve is that it tapers off more slowly than a normal curve, so some people get much higher scores than a normal curve would produce. If we assume that the distributions of ability and effort are closer to the normal curve, some process must be boosting the top scores, magnifying differences of ability and effort into large differences in outcomes.⁵⁰

Comparing the last two tables suggests that something of the kind is operating here. The quality of top twenty opinions, which presumably reflects the ability and effort of the judges fairly accurately, follows a normal distribution. But the overall number of citations does not, which indicates that for top judges, differences in citations for non-top twenty opinions are disproportionate to differences in ability and effort. In other words, there is a feedback loop for citations of less significant cases but not for citations of the most significant ones. We could conjecture that significant opinions are cited purely on the basis of their quality, but that less significant opinions are cited partly because the judge who wrote them is well known.

We might also consider the figures for law review citations to judges in Table 4.

TABLE 4
LAW REVIEW CITATIONS⁵¹

NUMBER OF LAW REVIEW CITATIONS	NUMBER OF JUDGES
0-199	14
200-399	44
400-599	29
600-799	10
800-999	0
1000-1199	1

49. *Id.* at 107.

50. This feature applies to a broader set of distributions than power curves. Curves having leptokurtosis have steeper peaks and fatter tails than the normal curve, but they have small shoulders—in other words, most cases are either near the average or spread out, with fewer intermediate cases than one would expect. Such distributions are said to be characteristic of situations involving actors with bounded rationality. See BRYAN D. JONES, POLITICS AND THE ARCHITECTURE OF CHOICE: BOUNDED RATIONALITY AND GOVERNANCE 164-84 (2001).

51. Figures collated from Choi & Gulati, *Empirical Ranking*, *supra* note 2, at 94-99 tbl.D.

With the possible exception of the top judge (once again, Posner) who seems to be something of an outlier, this distribution does not show strong signs of a superstar effect. Unlike invocations of judges in other opinions, citations to judges in law reviews seem to be distributed among a broader group of judges, with little evidence of the superstar phenomenon. This is all the more interesting because opinions in casebooks show strong signs of the superstar phenomenon,⁵² so law professors appear to use different criteria in deciding which cases to use in these two different settings.

Of the various citations figures, which is most relevant for considering potential Supreme Court appointments?⁵³ The answer is probably the outside circuit citations to the judge's top twenty opinions. The most important thing a Justice does is write majority opinions. Justices write fewer than twenty such opinions per year—more like ten these days. The tournament covers three years, so the “top twenty” count measures approximately the best seven opinions a judge can write in a year, which is pretty close to the Supreme Court workload. Citation figures for less significant cases tell less about Supreme Court performance. When you get down to the thirtieth least significant opinion a judge can write in a given year, some judges may be much better than others, but this is irrelevant since Supreme Court Justices never write that many majority opinions. In other words, what we care about is the quality of the work the judge does when he or she is really focused on a case and trying hard, not the quality of his or her work on less significant cases.⁵⁴

Another reason to favor the “top twenty” index is that it is more or less a normal distribution, displaying few signs of the superstar phenomenon or other effects that might amplify or distort differences between judges. Invocations are the worst index by this standard, but total citations also show more signs of nonlinear effects than top twenty citations. Law review citations also have some appeal in terms of the absence of apparent nonlinear distortions, provided we are willing to view citation decisions by academics as unbiased and accurate.

In the end, however, the choice of citation basis probably is not critical if we are only looking for a rough measure of opinion quality.

52. See Gulati & Sanchez, *supra* note 30.

53. Choi and Gulati also collect statistics on citations by the Supreme Court. The problem is that the numbers are small, ranging from 0 to 16, with almost all judges in a small range between 4 and 8. Thus, the numbers could represent mostly random variation. The only judges whose citation counts are listed as statistically significant (at the 95% level) are Posner, Easterbrook, and a group of judges who received zero citations. It would be interesting to examine the Posner/Easterbrook citations to get a better sense of when they are cited. See Choi & Gulati, *Empirical Ranking*, *supra* note 2, at 94-99 tbl.D.

54. The judge who tops the list based on this standard is Sandra Lynch. *Id.* at 100 tbl.E. Lynch replaced Stephen Breyer on the First Circuit.

As Choi and Gulati point out, the correlation between citation indices is very high after correcting for nonlinearities in some of the indices.⁵⁵ But if we are looking to pick a winner in a tournament, the choice of index may be more important.

IV. THE RELEVANCE OF OBJECTIVE MEASURES OF PERFORMANCE

Given what we can measure about past judicial performance, are these measurements relevant in evaluating possible Supreme Court nominees? Of the available measures, citation figures seem to be the best indication of professional performance for judges. But of course, they are only an imperfect measure. Indeed, among top judges, differences between citation rates might not even be statistically significant, so picking judges purely on the basis of numerical scores might not be sensible.

There is another reason why the President might give only limited weight to citation figures. I suggested earlier that the most relevant citation figures are probably the number of citations to a judge's twenty most influential opinions. The reason is that this is comparable to the number of majority opinions for a Supreme Court Justice in any given year. But the most heavily cited opinions by a given judge may not be a random sample of different legal issues.

Based on a recent study of citations to Supreme Court opinions,⁵⁶ it appears that the judicial opinions most heavily cited by other judges may tend to involve procedural and other technical issues.⁵⁷ For example, in the 1984 Term, the five opinions most cited by other courts involved procedural issues (three cases), ERISA, and an erroneous jury instruction on municipal liability.⁵⁸ In contrast, law reviews tend to favor citations to cases involving important social issues. Consequently, only a weak correlation exists between law review citations of Supreme Court cases and citations in law.⁵⁹

This might suggest that the President ought to focus instead on law review citations as better measuring performance in cases involving major social issues. The problem is that these citations arguably may be less reliable as an indicator of quality. Academics do not have the same incentives that judges have to focus their citations on the strongest opinions. They may equally well cite all of the opinions that bear on a particular topic, such as the legality of a specific form of abortion, or they may choose to write about the opinions with

55. See *id.* at 74. They correct for nonlinearities such as the superstar effect by using the logarithm of citation counts rather than the raw counts. *Id.* at 72.

56. See Farber, *supra* note 27.

57. See *id.* at 870.

58. See *id.*

59. See *id.* at 871, 873.

which they most disagree. Moreover, the President might reasonably fear that the political views of law professors would influence their citation practices, which would make law review citations particularly suspect if the President does not happen to share that political orientation. So the President might well find the “top twenty” citation rate to be a more reliable figure.

What this means is that most reliable citation figures probably best measure how well a judge does with technical legal problems. This is surely not an irrelevant consideration for a President. Presumably, it is in the country’s interest to have technical legal issues decided correctly and with the most useful possible guidance to the lower courts. But a President might reasonably be more concerned with how well a judge will handle cases involving major social issues such as abortion, affirmative action, or antiterrorism issues. Citation counts may not give a good indication of the quality of those opinions.

The President might take either of two views about how cases involving major social issues relate to more technical legal cases. On the one hand, he might believe that cases involving major social issues require only the same set of problem-solving and analytical skills involved in more technical cases. Even so, he might still have doubts about relying too heavily on the technical cases as gauges of performance in “hot” cases. He may be concerned about whether a judge who has very strong skills (as shown by citation levels) is able to exclude ideological factors in more political cases. Thus, he would want to look at those cases specifically to see if the judge is able to solve really contentious legal issues through a dispassionate application of legal skills.

On the other hand, the President may believe that decisions in “hot” cases involve something more than the application of technical legal skills—that personal values or empathy or statesmanship are especially important in these cases. Citation rates in cases with more legalistic issues will only weakly reflect the presence of these values.⁶⁰ Whichever view the President takes about the relationship between technical cases and “hot” cases, measuring the quality of opinions in the more technical cases would give him only limited guidance.

So far, I have been assuming that the President finds it irrelevant how the judge will vote on particular legal issues but is only looking for general indicators that a Justice will have the character and skill to make good decisions in hard cases. But it is not clear that a con-

60. This may be part of the explanation for why individuals with federal circuit experience have not historically tended to rank as “great” Justices, as observed by Steven Goldberg, *Federal Judges and the Heisman Trophy*, 32 FLA. ST. U. L. REV. 1237 (2005).

scientious President should take that view.⁶¹ There are two reasons why the President might give weight to his view of how the Justice will vote on key issues.

First, if the President has sufficient faith in his own views of the constitutional merits of these issues, he might view a judge's ability to reach the "right answers" to be a critical test of judicial ability and character. A judge who keeps getting the wrong answers must be doing something wrong, even if the President does not feel confident of his ability to determine the exact nature of the deficiency.

Second, the President might view the outcomes in those cases as having a social significance that outweighs any question of technical legal correctness. He may believe that abortion is a form of mass murder or, alternatively, that access to abortion is critical in order for women to be equal citizens. It might be wrong for a judge to base a decision entirely on her views about such social and moral considerations. But the President is not a judge and would seem to have the discretion to nominate the judge whom he thinks will be, all things considered, best for the nation's welfare.

The point is simple: If the President simply wants to make the best possible appointment, general measures of professional competence are not likely to be the exclusive factor in the decision. There might, however, be some collateral reasons for excluding or at least minimizing other factors.

To begin with, the President might believe that introducing other factors undermines confidence in the courts by making them appear politicized. This is certainly a concern, and one that has been frequently voiced. But nomination controversies do not seem to have harmed the courts' reputation to date, perhaps because few people pay attention to the appointments process.

Alternatively, the President might believe that basing his selection on apparently objective measures will lead to a smoother appointments process, with less unseemly wrangling with Congress. This would surely be desirable. But it is another question whether it is realistic, given the country's current political polarization. And in any event, the tournament might not be successful in reducing partisanship. For instance, the putative winner of the tournament, Judge Posner, has taken controversial positions which would be offensive to important political factions. It seems unlikely that their opposition would be silenced by the strength of his professional credentials.

Or perhaps the President might believe, like Choi and Gulati, that using objective selection criteria would give lower court judges an in-

61. By conscientious, I mean a President who wants only to make the best appointment for the good of the country, excluding all partisan or personal considerations.

centive for high performance. Here, there are two questions. First, how strong would the incentive effect be? Only a few judges would be serious contenders for victory in the tournament, and the incentive effect would be strongest for them—but presumably the best performers are least in need of additional incentives. Second, is the gain in incentives for lower court judges worth making an otherwise less desirable Supreme Court appointment? Looking at the welfare of the country as a whole, the President may think that small changes in the performance of lower court judges are not valuable enough to be worth the sacrifice.

V. CONCLUSION

Choi and Gulati propose a tournament, whereby promotion to the Supreme Court would be based on objective measures of a circuit judge's performance. Although it is intriguing, such a tournament is not, at least at present, a plausible option. Of the three traits that Choi and Gulati seek to measure, one (productivity) is ambiguous as an indicator of quality, and another (independence) is hard to measure and ideologically charged. Citation counts—and in particular, Choi and Gulati's "top twenty" counts—provide a better metric. This measure of opinion quality is imperfect but seems relevant to the President's selection.

Yet, even if overall judicial performance could be perfectly measured, the best lower court judge might not be the best choice for the President. Presidents are legitimately concerned with other factors, including how the nominee might vote on current key issues and how the nominee is likely to approach future national controversies. The tournament idea puts too much weight on one dimension, albeit an important one.

Although the tournament may not be "ready for prime time," collecting and disseminating data on circuit judge performance could be very useful. For example, a website could post the best available measures of judicial quality.⁶² If it established a reputation for objectivity, such a website could be a valuable resource during Supreme Court vacancies. It would be up to the relevant actors—Presidents, Senators, press members, and the public—to decide on the relevance of the information. More complete information could well improve the quality of the debate.

Furthermore, such a website also could serve less dramatic, but still significant, secondary functions. It would allow law students to

62. The model could be Brian Leiter's influential website ranking law school faculties. See Brian Leiter, *Ranking of Law Faculty Quality 2003-04*, EDUCATIONAL QUALITY RANKINGS OF U.S. LAW SCHOOLS 2003-2004, at <http://www.utexas.edu/law/faculty/bleiter/rankings/> (Mar. 25, 2003).

make better-informed decisions about clerkships. Even without the prize of a Supreme Court nomination, simply posting the measures could also create some useful incentive effects. Finally, the website would be a valuable resource for scholars.

It would be terrific to reform the acrimonious Supreme Court appointment process, but this may be too much for mere law professors to achieve. If Choi and Gulati manage to have only a small incremental effect on judicial selection, that will be noteworthy. In the meantime, the data they have already collected will provide scholars with significant insights into patterns of influence among appellate judges.