

JUDGING MEASURES

David F. Levi[†] & Mitu Gulati*

I. INTRODUCTION

A recent article in the *New York Times* reported on research into the decisions of a set of sixteen immigration judges.¹ The sixteen judges had all received executive branch appointments after having been “vetted” for their political bona fides²—presumably their allegiance to the President’s policies. Under applicable civil service laws,³ such vetting was improper; selection was supposed to be non-partisan, based on merit without regard to the political affiliation and beliefs of the applicant.⁴ An examination of over one-hundred asylum decisions made by these judges, as compared to the norm in their jurisdictions, found that nine of them denied asylum claims at higher rates.⁵ The higher rates of asylum denial, the article suggested, followed from the inappropriate vetting.⁶ Commentators interviewed for the article suggested that these improperly selected judges should perhaps resign and seek reappointment or be moved to nonadjudicatory positions.⁷

The *Times* article is interesting for a number of reasons. First, the article reports on an attempt to systematically measure the effects on judicial decision-making of a selection system that took the political viewpoint of the judicial candidate into account as part of the selection process. Second, although the results of the study are suggestive, there is little attempt to think critically about whether the results necessarily point to the conclusions that are drawn—that is, that these judges are trying to push a political agenda to deny more asylum claims. It is equally plausible that these judges (especially the nine who denied claims at higher rates) had higher than normal levels of experience with asylum claims and, therefore, had a better than average understanding of which claimants were credible. Would the presence of this greater skepticism based upon experience—if genuine—negate the inference that these judges have been acting in a politically-biased fashion? Alternatively, perhaps these judges have different judicial philosophies about how to interpret the immigration laws (e.g., strictly

[†] Dean and Professor of Law, Duke University School of Law. Previously, Chief Judge of the United States District Court for the Eastern District of California. We are grateful to Jennifer Dominguez and Matthew Wolfe for their editorial comments and suggestions.

* Professor of Law, Duke University School of Law.

¹ See Charlie Savage, *Vetted Judges More Likely to Reject Asylum Bids*, N.Y. TIMES, Aug. 24, 2008, at A17, available at <http://www.nytimes.com/2008/08/24/washington/24judges.html>.

² See *id.*

³ 5 U.S.C. §§ 2301–02 (2005).

⁴ *Id.* See also Savage, *supra* note 1 (stating that the applicable federal civil service laws “stipulate that [judges] must be hired on the basis of merit under politically neutral criteria”). Additionally, an Inspector General’s Report for the Justice Department condemned the practice of such vetting. See U.S. DEP’T OF JUSTICE, AN INVESTIGATION OF ALLEGATIONS OF POLITICIZED HIRING BY MONICA GOODLING AND OTHER STAFF IN THE OFFICE OF THE ATTORNEY GENERAL 14–15 (2008), available at www.usdoj.gov/oig/special/s0807/final.pdf.

⁵ See Savage, *supra* note 1.

⁶ See *id.*

⁷ *Id.*

and narrowly as opposed to loosely). Would this difference in interpretive philosophies count as political bias?⁸ Third, consider whether the use of these simple measures (grant asylum claim = liberal; reject asylum claim = conservative), and the accompanying public disclosure, could have any effect on the future decisions of these or other immigration judges. As best we can tell, none of the sixteen judges was willing to be interviewed for the *Times* article, which is not surprising: judges usually explain their decisions in written or oral opinions, not in later comments and justifications offered to journalists or academics.

This Essay attempts to address some of the questions and issues raised by the *Times* article. We advocate for the value of measuring judicial performance and the need to push the measurement project forward. But we also recognize that if the measurement project is to advance, changes to the research model must be made. The simple measures often used today, while necessary, cannot be relied on exclusively. In order to achieve a more reliable and useful measurement, judges must be involved in the process of arriving at the right characteristics to measure and the right ways to measure them. And, particularly if judges are involved in improving the quality of data collection and measurement, the inherent dangers in empirical analysis must not only be recognized but also navigated.

The Essay proceeds in three parts. Part II sets out the value of measuring judicial behavior. It then proposes collaboration between judges and researchers to identify appropriate characteristics of good judging, to determine what are fitting, measurable proxies for these characteristics, to collect the measurements, and to interpret the data. Part III reviews the existing body of scholarship that attempts to measure judicial behavior and exposes its inadequacy. As a way of filling the void and also beginning the dialogue between judges and scholars, Part IV argues for an initial focus on federal trial courts and suggests some important measures for the performance of federal district court judges.

II. UNANSWERED QUESTIONS

The optimal design of legal institutions is a central theme in legal scholarship.⁹ Consider the following ten questions, each of which deals with foundational design issues relating to judges, and each of which could be illuminated, if not fully answered, by empirical analysis:

⁸ The study addressed in the article did not compare the nine or sixteen judges to other judges at a similar early stage in their careers. *See id.* It is possible that new judges are less likely to grant asylum claims whether the judges are vetted or not. It also could be that the newness of their appointment may be combined with other possible factors; for example, if immigration judges are given a goal to aim at—a typical asylum grant rate—it may be that new judges are more determined to come in at the goal than more experienced judges, whether because new judges are less surefooted, more desirous of approval and promotion, or for other reasons.

⁹ *See infra* Part III(A).

- Should judges be selected by public elections, judicial elections, executive appointment, or by a civil service entrance exam?
- Should judicial salaries be set at, below, or above the market wage for equivalent lawyers, and should judges be permitted to earn outside income?
- How many levels of appeal are optimal and does the answer to that question differ as a function of the relevant substantive law?
- What is the value of racial, gender, age, religious, and other types of diversity in the judiciary and does this value depend on whether the particular court is a multimember court?
- What is the best method of promotion to the higher courts and selection to the highest court?
- Should judges be given life appointments or be required to retire at a certain age?
- What kind of staffing should be provided to judges, including whether judges should be provided with one or more law clerks and whether these law clerks should hold the position for long periods of time?
- Is there any measurable difference in performance between judges with a background in the private sector as compared to those with a background in the public sector?
- Do specialized courts produce better opinions and more consistent rulings than non-specialized courts?
- Should judges be subject to time limits to resolve pending matters?

These questions, and the many others that could be posed, can be answered empirically, in large part because of the rich possibilities for comparative analysis. Courts within the United States and around the world vary widely in terms of their institutional design. For instance, some systems have life appointments, while others have mandatory retirement. Some systems provide for the election of judges; others appoint them or use entrance exams, while still others employ psychological testing.¹⁰ Certain judicial systems might pay relatively high salaries (as compared to the private sector), while others may choose to pay comparatively low ones. Some judges are provided with law clerks and other assistants, whereas others are not. Further, the amount of legal experience these judicial assistants possess varies widely across systems. To determine which features work the best, one could compare the performances of judges and courts in the different systems. All other things being equal, if courts with mandatory retirement, high wages, and a civil service structure produce

¹⁰ See Mary L. Volcansek, *Appointing Judges the European Way*, 34 *FORDHAM URB. L.J.* 363, 374 (2007) (describing the Netherlands' use of entrance exams and psychological assessment).

better justice than courts with life employment, low wages, and judicial elections, then that might be an indicator of the superiority of the former set of features. From there, additional empirical tests could parse out the effects of the individual variables, the strength of these effects, and their interaction.

This is easier said than done. The first step in comparing any of the features of these divergent systems is to determine the quality of justice produced under each system. But how does one measure justice? The essence of judging is the exercise of judgment. Often there is no single, right outcome or approach. This is the case particularly at either end of the judicial spectrum: the highest court is likely to hear the most difficult cases presenting issues as to which there is no single “right” answer; the lowest courts are likely to be involved in case management and in the myriad discretionary decisions as to which there is no single “right” approach or finding.

That said, while it may not be easy to measure the resulting justice from a particular system feature, it is far from impossible. Judges are in the business of making decisions—usually, public decisions, with stated reasons, and their discretion is not unlimited. Similar issues or disputes recur in a variety of courts and the resolutions of these disputes can be compared and evaluated. Moreover, certain design features are likely to lead to a lower or higher quality of justice as those features become more extreme or vary radically from the norm. For example, judges in one jurisdiction who are severely overworked or understaffed probably perform less well according to certain measures—time to decision and adequate explanations for decisions—than those judges with fewer cases or more resources. And as with other crafts, there are artifacts—decisions or opinions—that can be expertly evaluated for the quality of decision-making.

To look at the scholarship which addresses questions relating to the optimal design of legal institutions would suggest that the measurement problem is largely insurmountable. There is scant empirical research tackling the basic question of how to compare the justice produced by courts with different institutional features. Relatedly, at least in the context of the United States, little is written about institutional reform of basic court characteristics. For example, one rarely comes across serious discussion of whether a civil service judicial system should be used in the United States, either in the state or in the federal system, despite the prevalence of such systems in Europe and Asia. Perhaps the dearth of empirical research on the quality of individual judges or courts reflects the view that exact measurement of judicial quality is not possible.¹¹ But while perfect measurement of judicial performance is unlikely, there are few if any jobs where the quality of employee performance can be measured perfectly. For example, the quality of services provided by a doctor or lawyer eludes simple measurement. In each of these jobs, there is a heavy dose of judgment required of the service provider—and it is the exercise of that good judgment in

¹¹ Cf. William P. Marshall, *Be Careful What You Wish For: The Problems With Using Empirical Rankings to Select Supreme Court Justices*, 78 S. CAL. L. REV. 119 (2004) (arguing that using empirical methods to measure judicial excellence is fraught with difficulty and may actually provide a cover for selecting justices on the basis of ideology alone).

identifying problems and providing solutions that is most valued. But, as with judges, good judgment is difficult to measure in a quantifiable fashion.

Despite the measurement problems for doctors, lawyers, and many other professions and crafts, employers and customers figure out ways to measure their performances all the time. Employers and customers of these service providers have no other choice but to figure out ways to measure performance, even if the measure is crude. Otherwise, there would be no method for determining selection and compensation. For example, lawyers are often paid in terms of the number of hours worked and are evaluated by employers on those terms as well. The measure is both crude and open to manipulation. But at some level, the employer knows, for example, that the law firm associate who billed 2,500 hours probably is working harder than the one who billed a mere 1,600 hours. With senior litigators or transactional lawyers, one might use other measures such as rates of victory, settlement, or deal completion. Prosecutors might be measured by the number of indictments, the number of closed cases, or their conviction rate within a certain time period. While these are imperfect measures and subject to manipulation and abuse by both employer and employee, they are used because even imperfect measures can provide useful information and can shape the way employees behave.

Judges and courts should be no different. Imperfect comparative information should have utility in evaluating the judiciary and determining how best to reform it. Note that in comparison to most other service providers—who often work in isolation with clients on individualized problems and do not always produce public work-product—judges should be easier to measure because they produce public outcomes. That is, they generate decisions and orders in cases where the arguments being made are documented and so is the rationale for the decision or order. Because individual judges in different jurisdictions are often tackling very similar types of disputes and are producing the same types of outputs, there should be ways of looking at those outputs and comparing them across jurisdictions, with their differing institutional variables.

One of the curious aspects of some of the most obvious output measures is that they are not linear and only work best at the extremes. But this does not render them useless. For example, if the judges in one jurisdiction are consistently writing ten-page explanations for their decisions and the judges in another jurisdiction are writing one-sentence explanations for the same number and type of cases, that suggests that something different is going on in the two court systems and that it might be useful to inquire further. On the other hand, one would not conclude that a fifty-page opinion is better than a twenty-five-page opinion; it well might be worse. When opinions are drafted by law clerks who include information in the opinion that the judge may or may not want, a longer opinion could be one that has escaped serious editing by the judge—not a sign of a better judicial output. In reaching some conclusion about which is “better,” the solution might be to get a sense of the normal length of opinion for the type of case at issue. If a judge or a court deviates from this norm, it might be an indication of something gone awry or something working particularly well, and thus may be cause for further study and evaluation. For a similar example of the problem of non-linearity, consider reversal rates as between judges or courts. At

the extremes, comparing a judge who is rarely reversed with a judge who is frequently reversed, the reversal rate might well say something about the quality of the judge or the court on which the judge sits or by which the judge's decisions are reviewed. However, one would not necessarily draw the conclusion that a judge who is reversed more often than another judge is a less effective judge. To the contrary, a judge who is rarely reversed may be knowingly or unknowingly manipulating the system to avoid review by placing the ground for decision on areas of discretion whether or not this is the principal ground of decision. Or the never-reversed judge may be one who is overly cautious and never exercises creativity or attempts even the most modest law reform, qualities that could be associated with judicial excellence. As with opinion length, a calculation of the normal reversal rate may help to evaluate those courts or judges whose reversal rate seriously deviates from the mean. Other output criteria can follow a similar pattern of comparison with the norm. For example, while the judge with the quickest time to decision might not be categorized as the "best," significant deviation from the mean should be of interest.

Output measures of a different sort may provide for comparisons without reference to a norm. When there is a set of cases in which the facts are relatively similar, the decisions of judges can be compared. Sentencing statistics, for example, can yield data on whether judges treat litigants of different races and genders differently, and whether they are influenced by factors that the sentencing law might not deem relevant or appropriate.¹²

There are input measures that could be used in the measurement project as well. One could look at the number of business contracts that specify that the parties have agreed to litigate disputes in front of a particular court. Or one could examine the substantive law that is chosen to govern contracts. Where they have a choice, contract lawyers will likely choose a system that they perceive is fair, subjects them to minimal delay, and possesses good judges. If the law and courts of New York State are consistently chosen in sophisticated corporate deals, particularly those in which the parties are not located in New York, this could be an indication of superior quality in that law and that court. Perhaps it means that the New York legislature is more attentive to legal issues and acts quickly whenever there are problems.¹³ Or perhaps it means that the judges in New York are more familiar with corporate transactions and, thus, their decisions are fairer and more predictable. Either way, the data on the rates at which different laws are chosen in contracts allows for a plausible inference about law or court

¹² See, e.g., Susan Katzenelson & Kyle Conley, *Guideline Sentences in the Ninth Circuit*, in FINAL REPORT NINTH CIRCUIT TASK FORCE ON RACIAL, RELIGIOUS, & ETHICAL FAIRNESS 54-56 (1997); Frank O. Bowman, III & Michael Heise, *Quiet Rebellion II: An Empirical Analysis of Declining Federal Drug Sentences Including Data from the District Level*, 87 IOWA L. REV. 477 (2002); Frank O. Bowman, III & Michael Heise, *Quiet Rebellion? Explaining Nearly a Decade of Declining Federal Drug Sentences*, 86 IOWA L. REV. 1043 (2001).

¹³ Part of the appeal of empirical research is that scholars may be able to determine whether it is the substantive laws or the quality of the courts or judges that attracts parties to contracts to choose these forums.

quality.¹⁴ Similarly, one could estimate the relative numbers of cases going into private arbitration systems as opposed to the state-sponsored judicial system. Whether this number is comparatively high or low might provide insight into the perceived fairness, effectiveness or efficiency of arbitration versus court litigation. Alternatively, members of the bar, litigants, or members of the public could be surveyed to determine their knowledge and attitudes about the courts.¹⁵ At the least, such a survey measurement would provide a measure of public confidence in the official court system, the attainment of which is probably one goal of all court systems.¹⁶

In deciding which of these measures could be most useful in making a reasonable evaluation of the quality of justice, researchers would greatly benefit from the advice and input of individual judges and those agencies that serve the courts, such as the Federal Judicial Center.¹⁷ Making use of the expertise of the subjects of the study would enhance the usefulness and accuracy of the measures. Yet, in our experience, judges do not tend to react positively to attempts to measure relative court or judge performance.

That reaction is not surprising. Judges are constrained by their office from defending or explaining their performance. They may fear that political actors from other branches of government or, in the case of elected judges, potential rivals for office will seize upon such data and attempt to use it in ways that are unfair and misleading. Most of us do not appreciate our job performance being measured. And such critiques are especially galling when done by those who understand little about the job in question. Academics would not want the quality of their research critiqued and worse, ranked, by judges with little or no

¹⁴ Of course, this assumes that the parties are of equal bargaining power. Simply looking at where cases are filed or removed to only gives an indication of which court or judge is preferred by that party, and this could indicate a “positive” assessment (the party thinks that the court or judge is efficient) or it could be a “negative” assessment (the party thinks that the court or judge is biased in its favor). Alternative explanations for the choice of law are also possible, such as path dependence. Empirical examination starting out with the premise that choice of law indicates legal quality will generally try to eliminate these other explanations before drawing inferences from the differential rates of choice. See, e.g., Stefan Voigt, *Are International Merchants Stupid? Their Choice of Law Sheds Doubt on the Legal Origin Theory*, 5 J. EMPIRICAL LEGAL STUD. 1, 10-11 (2008); Theodore Eisenberg & Geoffrey P. Miller, *Ex Ante Choices of Law and Forum: An Empirical Analysis of Corporate Merger Agreements* 2-3 (N.Y. Univ. Law & Econ. Research Paper No. 06-31, 2006), available at <http://ssrn.com/abstract=918735>.

¹⁵ See, e.g., James L. Gibson et al., *On the Legitimacy of National High Courts*, 92 AM. POL. SCI. REV. 343 (1998) (analyzing public attitudes toward high courts in eighteen countries).

¹⁶ Arguably, public attitudes about the local court system are output rather than input measures. For purposes of our discussion, we refer to the formal outputs of courts—decisions, rulings, opinions—as outputs, and the attitudes, perceptions, and choices of those using the system as inputs. Each set provides an indication, albeit incomplete, of court quality.

¹⁷ The Federal Judicial Center is the education and research agency for the federal courts. Congress created the FJC in 1967 to promote improvements in judicial administration. Federal Judicial Center, <http://www.fjc.gov>.

experience with academic research.¹⁸ Judges are no different.¹⁹ Further, by focusing on that which can be measured, empirical evaluations may put pressure on judges, in turn, to change the focus of their efforts. This could lead to distortion and to a devaluing of those important judicial functions and efforts for which no “credit” is awarded. But these concerns, fair as some of them are, do not mean that the measurement project should not proceed at all.

Most importantly, the measurement and evaluation is going to move ahead whether judges like it or not. Outside the legal academy, there is already a literature that purports to measure and compare the performances of court actors.²⁰ Some of this is special pleading by interest groups in the context of judicial elections. Much of it is serious social science research. And that literature is only growing. For the most part though, that literature has tackled but a handful of the institutional design questions mentioned at the start of this Essay.

The solution, we suggest, is for researchers and judges to cooperate. With that cooperation, better measures of judicial behavior can be devised; more accurate data collection measures can be implemented; and a focus on collecting the type of data that can lead to improved institutional design can be brought to bear. Judges probably understand what they do better than academics. Further, they have research resources and access to data. In the U.S. context, there are at least two large research institutions with available resources—the Federal Judicial Center and the National Center for State Courts²¹—that could collect data and process these improved measures. Moreover, if judges were to engage with the measurement debate, it is probable that the types of information that would be generated would help not only researchers, but also the judges themselves.²² Finally, if judges were involved, the limitations of the measures, and their possible distorting effects on future judicial behavior, would be identified, better understood, and perhaps overcome.

Getting judges interested in the measurement task could also improve researchers’ access to information. Given the public nature of most judicial

¹⁸ In fact, the most prominent ranking of law school quality relies heavily on judicial evaluation of particular law schools.

¹⁹ In our experience, many judges deeply resent the “attitudinal” model. If turnabout is fair play, it might be amusing for some judge to study the political registration of academics and the positions they take in their research and scholarship. Perhaps some judge will undertake this study and develop a “meta” attitudinal model (those academics who subscribe to the attitudinal model of judging tend to be associated with a particular political registration or a degree of partisanship). Since academics also seek appointment to government and judicial posts, we might be able to develop measures of their independence from such concerns.

²⁰ E.g., Edward L. Glaeser et al., *What Do Prosecutors Maximize? An Analysis of the Federalization of Drug Crimes*, 2 AM. L. & ECON. REV. 259 (2000).

²¹ In addition to the FJC, discussed *supra* note 17, the National Center for State Courts is an independent, nonprofit organization. Its mission is to improve the administration of justice through leadership and service to state courts, and courts around the world. National Center for State Courts, <http://www.ncsconline.org>.

²² See *infra* Part IV.

output, data on judges is plentiful. But there are aspects of the judging job that are secret, such as how and when appellate judges prepare opinions in routine cases, whether bench memoranda are prepared by each judge's staff, or whether there is one memorandum circulated among the chambers. Without judicial involvement, researchers often have to make assumptions about institutional practices—for example, regarding the levels of delegation to law clerks—based on anecdotal evidence. That is hardly an ideal starting point. If judges were to cooperate in the measurement process, they may decide it would be useful to help collect the data and then collaborate with researchers on the analysis of some of this information for everyone's benefit.

Our goal is to urge a conversation between judges and the academics who study them; that is, we desire to persuade the researchers and, more importantly, the subjects of their study, that there is value to engaging in discussion and debate over measuring the judiciary. As a precursor—to show both how extensive the measurement literature is and how impoverished it is—Part III describes various measures that have been used already. Then, in Part IV, using the context of the U.S. federal trial courts, an area that has received scant attention in the empirical literature on judging, we make suggestions as to the types of measures that might be useful and that realistically could be attempted.

III. A PANOPLY OF MEASURES

Three bodies of literature dominated by scholars from economics and political science with a handful of contributions from legal scholars—and none, as best we can tell, in conversation with the other—bear upon our inquiry.

The first body of literature, primarily the product of scholars in development economics and finance, is interested in the question of how to set up institutions to spur economic growth (the focus often being developing or transition economies). And legal institutions—in particular, courts and laws—are key among the institutions studied.

The second body of literature is U.S.-based and comes largely from scholars in economics and political science. The objective is to use the variation among the judicial systems in the fifty states to determine whether electoral, appointment, or merit selection systems work best.

The third body of literature, primarily from political science and, more recently, the legal academy, asks how best to measure some variant of judicial quality, greatness, or reputation. Unlike the economic development literature on legal institutions, which looks at cross-country comparisons, or the elections versus non-election literature that looks at cross-state comparisons within the United States, this literature tends to have a localized focus, often on judges within a single system—such as the U.S. or the Australian appeals courts.

Below we describe in general terms some of the measures used in the three bodies of literature. We have two goals: first, to show that, largely outside of the legal academy, there already exists a large literature on the measurement of legal institution quality; and, second, given the inadequacies of the existing literature, to suggest that there is room for legal scholars and judges to contribute to the measurement literature by offering improvements to the possible measures.

A. Legal Institutions and Economic Growth

The intellectual history of the various iterations of the idea that law and legal institutions can affect economic growth and well-being is long and dates back to Max Weber, if not before.²³ Serious attempts to measure and compare various legal institutions and systems and to relate those measures to growth, however, are of recent vintage. And those attempts are the product of the modern development economics and finance literature. The theoretical foundations come out of the work of Douglass North,²⁴ Mancur Olson,²⁵ and Oliver Williamson.²⁶ Economic activity occurs most effectively in contexts where exchange is easy. Legal institutions serve to enable easy exchange in a variety of ways—for example, by providing clarity of property rights and flexibility in protecting against opportunism problems.²⁷ Building on those theoretical foundations, scholars since have looked at a host of various proxies for the quality of legal institutions to see whether they relate to economic growth.²⁸

1. Broad Measures

Among the proxies used initially were levels of political instability and the nature of the political system, with the rationale that political instability or an authoritarian political system might reduce the protection given to individual property rights, thereby increasing the risks of engaging in economic activity.²⁹

²³ Kevin E. Davis & Michael J. Trebilcock, *The Relationship Between Law and Development: Optimists versus Skeptics*, 57 AM. J. COMP. L. (forthcoming 2009) (manuscript at 40), available at <http://ssrn.com/abstract=1124045>; see also David Kennedy, *The "Rule of Law," Political Choices, and Development Common Sense*, in THE NEW LAW AND ECONOMIC DEVELOPMENT: A CRITICAL APPRAISAL (David M. Trubek & Alvaro Santos eds., 2007).

²⁴ See DOUGLASS C. NORTH, INSTITUTIONS, INSTITUTIONAL CHANGE AND ECONOMIC PERFORMANCE 8 (Cambridge Univ. Press 1990).

²⁵ See Mancur Olson, *Big Bills Left on the Sidewalk: Why Some Nations are Rich, and Others Poor*, 10 J. ECON. PERSPECTIVES 3, 20 (1996), available at http://www.fbird.com/assets/MancurOlson_on_Transition_Econ_732003152238.pdf.

²⁶ See ORGANIZATION THEORY: FROM CHESTER BARNARD TO THE PRESENT AND BEYOND 182 (Oliver E. Williamson ed., Oxford Univ. Press 1995).

²⁷ The idea that robust property rights are key to ensuring economic growth is often associated with the work of De Soto. See generally HERNANDO DE SOTO, THE MYSTERY OF CAPITAL: WHY CAPITALISM TRIUMPHS IN THE WEST AND FAILS EVERYWHERE ELSE (Basic Books 2000).

²⁸ E.g., Gerald W. Scully, *The Institutional Framework and Economic Development*, 96 J. POL. ECON. 652, 652 (1988) (estimating, using a cross-country sample of 115 sovereigns, that sovereigns with good institutions grow roughly three times as fast as those lacking them).

²⁹ Considerable evidence was found to suggest that political instability did hamper growth. See, e.g., Alberto Alesina & Roberto Perotti, *Income Distribution, Political Instability, and Investment*, 40 EUR. ECON. REV. 1203, 1225 (1996); Robert J. Barro & Jong-Wha Lee, *Sources of Economic Growth*, 40 CARNEGIE-ROCHESTER CONF. SERIES ON PUB. POL'Y 1, 1 (1994). The evidence that dictatorships reduced growth as posited, see Douglass C. North & Barry R. Weingast, *Constitutions and Commitment: The Evolution of Institutions Governing Public Choice in Seventeenth-Century England*, 49 J. ECON. HIST. 803, 807-08 (1989), was not as strong. Alberto Alesina & Roberto

Some scholars also looked at measures of country risk generated by intermediaries that were sold to private firms evaluating investment opportunities. These country risk measures tended to incorporate the risks to external capital of investing in particular economies.³⁰ Other researchers have looked to proxies for the “rule of law,” such as confidence in the judiciary or corruption rankings—measured by surveys of business firms, legal practitioners and even legal scholars.³¹ Across a variety of these studies, common law jurisdictions have typically outperformed their civil law cousins.³²

In recent years, these broad survey-based measures have received special prominence in part because of the role the World Bank has played both in encouraging this research and in putting the insights from such research at the center of its economic development strategies.³³ These proxies for legal quality are rough, measuring both a lot more and a lot less than legal quality. For example, the measures of country risk were often focused primarily on issues that mattered to foreign investors or, still problematic, large business firms which might not translate necessarily into what was relevant for the majority of transactors in the local economy.

Perotti, *The Political Economy of Growth: A Critical Survey of the Recent Literature*, 8 WORLD BANK ECON. REV. 351, 353 (1994).

³⁰ See, e.g., Stephen Knack & Philip Keefer, *Institutions and Economic Performance: Cross-Country Tests Using Alternative Institutional Measures*, 7 ECON. & POL. 207, 223 (1995); Paulo Mauro, *Corruption and Growth*, 110 Q.J. ECON. 681, 683-84 (1995); see also Christopher Clague et al., *Property and Contract Rights in Autocracies and Democracies*, 1 J. ECON. GROWTH 243, 254 (1996) (looking to the fraction of “contract-intensive money”—that is, money in the economy that depends on third party contract or property enforcement).

³¹ E.g., ROBERT J. BARRO & XAVIER SALA-I-MARTIN, *ECONOMIC GROWTH* (2d ed. 2003); WORLD BANK, *WORLD DEVELOPMENT REPORT 2002: BUILDING INSTITUTIONS FOR MARKETS* 129 (2002), available at <http://www.worldbank.org/wdr/2001/fulltext/ch6.pdf>; WORLD BANK, *WORLD DEVELOPMENT REPORT 2000/2001: ATTACKING POVERTY: OPPORTUNITY, EMPOWERMENT, AND SECURITY* 8 (2000), available at <http://siteresources.worldbank.org/INTPOVERTY/Resources/WDR/overview.pdf>; WORLD BANK, *DOING BUSINESS IN 2004: UNDERSTANDING REGULATION* 65 (2004), available at <http://www.doingbusiness.org/Documents/DB2004-full-report.pdf>; Witold J. Henisz, *The Institutional Environment for Infrastructure Investment*, 11 INDUS. & CORP. CHANGE 355, 356 (2002), available at <http://www-management.wharton.upenn.edu/henisz/papers/iei.pdf>; Joseph L. Staats et al., *Measuring Judicial Performance in Latin America*, 47 LATIN AM. POL. & SOC’Y 77 (2005); see also Kevin E. Davis & Michael B. Kruse, *Taking the Measure of Law: The Case of the Doing Business Project*, 32 L. & SOC. INQUIRY 1095, 1104 (2007), available at http://www.doingbusiness.org/documents/Taking_the_Measure_of_Law.pdf.

³² E.g., Daniel Kaufman et al., *Governance Matters VI: Governance Indicators for 1996–2006* (World Bank Pol’y Research Working Paper No. 4280, 2007), available at <http://ssrn.com/abstract=999979>; Stijn Claessens & Luc Laeven, *Financial Development, Property Rights, and Growth*, 58 J. FIN. 2401, 2402 (2003); Thorsten Beck et al., *Law and Finance: Why Does Legal Origin Matter?*, 31 J. COMP. ECON. 653, 654-56 (2003).

³³ See Kevin E. Davis, *What Can the Rule of Law Variable Tell Us About Rule of Law Reforms?*, 26 MICH. J. INT’L L. 141 (2004); see also Doing Business Blog, The World Bank Group, <http://blog.doingbusiness.org> (last visited Sept. 10, 2008).

2. Examining Specific Laws

Given the criticisms of the broad indicators, scholars began measuring more specific features of the legal system. Focus has been primarily on the levels of protections given to different types of property rights—protections to property rights being a key element of the theoretical foundations for this research. And this research yielded promising results. The protections provided, for example, to intellectual property rights, were found to positively correlate with various indicia of economic growth.³⁴ More recently, the robustness of protections given to minority investor rights have been measured so as to determine the relationship between those protections and the growth of capital markets.³⁵ Other areas of substantive law comparisons, as a way of explaining differential rates of capital markets growth, include bankruptcy laws (that could be compared in terms of their effectiveness in eliminating inefficient firms) and secured credit laws (that allow for collateral to be used effectively as a basis for financings). Although it is hard to argue with the premise that substantive law can make a difference in how legal institutions affect economic growth, measuring the effects of substantive law alone also is inadequate.

Among the problems with correlating these substantive laws to growth rates is that causality is difficult to determine; the correlations do not reveal whether robust legal rights caused the economic growth or resulted from it. In addition, as a theoretical matter, the examination of rights alone misses half the legal puzzle. Rights, on their own, do not necessarily translate into economic efficiency. Indeed, too many rights can sometimes bring a system to a standstill, as, for example, in the case of minority bondholders and debt restructurings in bonds governed by U.S. law.³⁶ An effective legal system is one where the judges are able to understand and balance the hold-out risk posed by minority creditors

³⁴ David M. Gould & William C. Gruben, *The Role of Intellectual Property Rights in Economic Growth*, 48 J. DEV. ECON. 323, 323 (1996) (using cross-country regressions to analyze the impact of intellectual property rights on growth and finding evidence that stronger rights correlate with higher growth rates); EDWIN MANSFIELD, *INTELLECTUAL PROPERTY PROTECTION, FOREIGN DIRECT INVESTMENT, AND TECHNOLOGY TRANSFER* 1 (1994) (finding an empirical relationship between strength of intellectual property rights and the ability to attract foreign investment); Edwin Mansfield, *Intellectual Property Protection, Direct Investment, and Technology Transfer: Germany, Japan, and the United States* 1 (Int'l Fin. Corp., Discussion Paper No. 27, 1995) (same findings). But cf. Elhanan Helpman, *Innovation, Imitation, and Intellectual Property Rights*, 61 ECONOMETRICA 1247, 1274-75 (1993) (suggesting that while strong intellectual property rights might help industrialized countries, they may not benefit developing countries).

³⁵ Rafael La Porta et al., *Legal Determinants of External Finance*, 52 J. FIN. 1131, 1149 (1997).

³⁶ Minority bondholders in the United States have the right to veto any restructuring of public debt that reduces the value of their claim. This right gives minority creditors power and would likely get measured by one of the above-mentioned studies as a high level of substantive rights. But taken on its own, such a veto right increases the costs of restructuring and reduces overall economic efficiency. Mark J. Roe, *Voting Prohibition in Bond Workouts*, 97 YALE L.J. 232, 279 (1987). These minority rights, particularly in sovereign bond restructurings, can cause delays in implementing workouts. See Lee C. Buchheit & G. Mitu Gulati, *Sovereign Bonds and the Collective Will*, 51 EMORY L.J. 1317, 1327 (2002).

against the risk of opportunism by majority creditors. Perhaps even more important to economic growth, one can tell a similar story with respect to intellectual property rights: grant too many of them and the result can be to deter innovation.³⁷ Ideally, there needs to be a balancing between the grant of rights and a flexibility that protects against inefficient holdup by right holders. One way to have that balancing occur is through experienced, capable judges, who have adequate discretion, and who can recognize when to apply the right strictly and when not to.

Although the empirical focus has been primarily on measuring private laws such as corporate, intellectual property, and contract law, there is a sizeable theoretical literature that suggests that public laws matter enormously as well.³⁸ Public laws arguably represent a bargain among the different segments of the population, with powerful interests promising not to expropriate from or disadvantage minority interests. In this vein, there have been attempts to measure the stability of constitutional bargains—for example, by looking at cross-country comparisons in the numbers of times basic constitutional provisions have been amended.³⁹ Again, the measures are rough—driving some of our constitutional law colleagues to distraction. The point though is that scholars are coming up with creative ways of not only theorizing the importance of different laws, but coming up with measures to test their ideas.

3. Legal Origins

Responding to some of the foregoing criticisms—particularly the concern about determining causality—the strand of literature that has garnered the most attention in recent years falls under the rubric of “legal origins.” The basic claim is the following: Jurisdictions with origins in the common law system, because of the considerable discretion that judges have in the law-making process as compared to civil law systems, are more likely to produce a legal system that induces economic growth. The suggestion also has been made that common law judges, because of the greater discretion and status given to them, are more likely to be independent of political control than their civil law counterparts and, therefore, more protective against encroachments by the state.⁴⁰ It also may be that judges in code-based versus precedent-based systems have differential abilities to commit to efficient rules, ex-ante, and that, because of the different hiring and promotion systems, the incentives of common law judges are more in the direction of allowing creativity, whereas those in civil law systems are to

³⁷ See generally JAMES BOYLE, *THE PUBLIC DOMAIN: ENCLOSING THE COMMONS OF THE MIND* (2008) (discussing how the public domain is hindered by the current system of copyright, trademark and patent laws).

³⁸ See, e.g., *id.*

³⁹ See, e.g., *INSTITUTIONS & PUBLIC LAW: COMPARATIVE APPROACHES* (Tom Ginsburg & Robert A. Kagan eds., 2005).

⁴⁰ Edward L. Glaeser & Andrei Shleifer, *Legal Origins*, 117 Q.J. ECON. 1193, 1221 (2002); Paul G. Mahoney, *The Common Law and Economic Growth: Hayek Might Be Right*, 30 J. LEGAL STUD. 503, 507 (2001).

push toward rule following.⁴¹ Finally, a number of articles link the substantive legal doctrines that come out of different systems (for example, the levels of rights of minority creditors) with the legal origins of the systems.⁴² Responding to the criticisms that examination of substantive rights alone is inadequate, more recent articles have looked in greater detail at not just substantive law, but also enforcement. For example, cross-country variation in the implementation of eviction for non payment of rent and recovery on a bounced check has been examined in detail, down to the number of stages in the process, the degrees of formality of process, and the time it takes to receive recovery. Variation across jurisdictions has then been related to variance in legal origin.⁴³ Again, the explanation given is often in terms of the different levels of discretion that common law and civil law judges have in law making—the discretion in common law systems, it is argued, allows more flexibility and informality and ultimately ease of transacting.⁴⁴ To the extent these articles delve deeper into the question of how judges in different systems are likely to behave as a function of the different institutional structures in which they operate, the articles get closer to inquiring into potential judicial performance and represent an advance on the research that solely examined the substantive content of laws across countries.

Even if one accepts the notion that legal origins can help determine whether judicial incentives are properly aligned and judges are provided with good information, there are gaps in the story. The legal origin line of scholarship, for the most part, has made minimal progress in terms of measuring actual judicial performance.⁴⁵ Good legal origins matter little if the judges running the system are of poor caliber. For example, among the virtues of the common law system that allows judges to take into account specific facts and craft efficient precedent for the future is that judges will produce detailed and nuanced explanations of their decisions that are widely distributed and serve to inform the legal

⁴¹ See, e.g., Luca Anderlini et al., *Statute Law or Case Law?* 4 (Ctr. for Econ. Studies & Ifo Inst. for Econ. Research, Working Paper Series No. 2358, 2008), available at <http://econ.lse.ac.uk/staff/lfelli/papers/Common.pdf>; J. Mark Ramseyer & Eric B. Rasmussen, *Judicial Independence in a Civil Law Regime: The Evidence from Japan*, 13 J.L. ECON. & ORG. 259 (1997).

⁴² See, e.g., Simeon Djankov et al., *Courts*, 118 Q.J. ECON. 453, 483-85 (2003) (reporting on the Lex Mundi project); Rafael La Porta et al., *The Quality of Government*, 15 J.L. ECON. & ORG. 222 (1999).

⁴³ See Djankov et al., *supra* note 42, at 457-59.

⁴⁴ See, e.g., *id.* at 460; Simon Johnson et al., *Tunneling*, 90 AM. ECON. REV. 22, 24 (2000); Edward L. Glaeser et al., *Do Institutions Cause Growth?* 9 J. ECON. GROWTH 271, 276 (2004); Rafael La Porta & F. López-de-Silanes, *Capital Markets and Legal Institutions*, in BEYOND THE WASHINGTON CONSENSUS: INSTITUTIONS MATTER (1998); Oliver D. Hart et al., *A New Bankruptcy Procedure That Uses Multiple Auctions*, 41 EUR. ECON. REV. 461 (1997); Juan C. Botero et al., *The Regulation of Labor*, 119 Q. J. ECON. 1339 (2006); Simeon Djankov et al., *The New Comparative Economics*, 31 J. COMP. ECON. 595, 620 (2003).

⁴⁵ See generally Gillian K. Hadfield, *The Quality of Law: Judicial Incentives, Legal Human Capital and the Evolution of Law* (Univ. of S. Cal. Ctr. in Law, Econ., & Org., Research Paper No. C07-3, 2007) (looking instead at the detailed institutional factors that influence the quality of law when judges have incentives to promote social welfare), available at <http://ssrn.com/abstract=967494>.

community and other judges faced with similar cases. Civil law courts, by contrast, are characterized as producing terse decisions with little explanation that are often not distributed widely.⁴⁶ But not all common law judges produce the types of published opinions envisioned. Incompetent judges will not provide the legal system with all the virtues that are posited for the common law; they may well do the reverse. While the logic of basic claims in the legal origins scholarship may make sense, the weight attached to the superiority of common law systems over civil law ones seems disproportionate to reality. Civil law based systems are not doing that badly on the economic growth front. And many economies with common law origins are not doing that well. Finally, there are the puzzles posed by systems like India and China—two systems that in sheer size dwarf much of the rest of the developing world—where economic growth rates in recent years have been high despite there being in one case an essentially non-functioning legal system (India) and in the other a one-party autocracy (China).⁴⁷

4. Measuring Independence

Moving in the direction of looking more carefully at key institutional features of a court system that might suggest high quality, some of the economic growth scholars have attempted to measure judicial quality. Their focus has been almost exclusively on one aspect of judicial behavior, independence—implicitly, making the claim that independence is the key aspect of an effective judicial system that promotes economic growth.

In order to measure and compare independence levels, scholars have had to look to various institutional features that might serve as proxies for independence. Included among these proxies are the methods of nominating or appointing judges of the highest courts, judges' term lengths, the possibility of reappointment, the procedure for removing judges from office, judges' pay and protections against reduction of income, the accessibility of the court, the presence (or lack thereof) of a general rule allocating cases to specific judges, and publication requirements concerning the decisions of the court.⁴⁸ These

⁴⁶ *Id.*

⁴⁷ See John Armour & Priya P. Lele, *Law, Finance, and Politics: The Case of India* (Eur. Corp. Governance Inst. – Law, Working Paper No. 107/2008, 2008), available at <http://ssrn.com/abstract=1116608>; Franklin Allen et al., *Law, Finance, and Economic Growth in China*, 77 J. FIN. ECON. 57, 57 (2005), available at [http://finance.wharton.upenn.edu/~allenf/download/Vita/Finlaw_china_JFE_final_journal_form%20\(2\).pdf](http://finance.wharton.upenn.edu/~allenf/download/Vita/Finlaw_china_JFE_final_journal_form%20(2).pdf); cf. Paul Wachtel et al., *Institutional Development, Financial Deepening and Economic Growth: Evidence from China* 23 (Bank of Fin. Inst. for Econ. in Transition, Discussion Paper No. 12/2006, 2006), available at <http://ssrn.com/abstract=921624> (showing that political pluralism in Chinese provinces is associated with stronger economic growth).

⁴⁸ Lars P. Feld & Stefan Voigt, *Economic Growth and Judicial Independence: Cross Country Evidence Using a New Set of Indicators* 5-8 (Ctr. for Econ. Studies & Ifo Inst. for Econ. Research, Working Paper No. 906, 2003), available at <http://ssrn.com/abstract=395403>; Lars P. Feld & Stefan Voigt, *Making Judges Independent—Some Proposals Regarding the Judiciary* 8-9 (Ctr. for

institutional features, the scholars argue, translate into judiciaries that can be expected to have more or less independence. Some of these assumptions, however, such as ones about high salaries translating into greater judicial independence or the relationship between the type of judicial appointment system and independence have not been rigorously tested.⁴⁹

The focus of the empirical work on legal institutions and economic growth, with but a handful of exceptions, has been on cross-country variation. That is, scholars have looked at the relationship among variables indicating the quality of a country's legal institutions—as noted, many of which were extremely rough—and variables measuring economic growth.⁵⁰ A smaller subset of economic research, also interested in measuring judicial independence, but focused on the U.S. courts, has examined the question of whether systems using judicial elections produce greater levels of independence than those using appointments or merit selection.⁵¹ Examining the variation within a single economic system like the U.S. has benefits in that a number of differences across countries that might have confounded the results from cross-country comparisons—such as differential trade policies, macroeconomic policies, central bank independence levels, and participation of the state in investment⁵²—can be eliminated from the analysis. On the downside, the variation within a system tends to be lower than that in the international context and sometimes can be illusory.

A related literature that looks almost exclusively at one aspect of judicial independence focuses on whether judges are biased towards the policies and interests of the political parties they were affiliated with prior to becoming

Econ. Studies & Ifo Inst. for Econ. Research, Working Paper No. 1260, 2004), *available at* <http://ssrn.com/abstract=597721>.

Another innovative technique used to estimate judicial independence has evaluated the stock market reactions to laws providing greater judicial independence (the proxies for independence being protection for the judiciary from executive interference in termination and salary decisions). See Daniel Klorman & Paul G. Mahoney, *The Value of Judicial Independence: Evidence from 18th Century England* (Univ of Va. Law & Econ. Research Paper No. 03-12, Univ. S. Cal. Law & Pub. Pol'y Research Paper No. 04-2, Univ. S. Cal. Law Sch. Olin Research Paper No. 04-2, 2004), *available at* <http://ssrn.com/abstract=495642>. Elsewhere, in the Japanese context, for example, decisions about the transfers of judges have been correlated with whether those judges ruled in favor or against the ruling party to examine whether judicial independence was being punished. See Ramseyer & Rasmussen, *supra* note 41.

⁴⁹ For example, on salaries, see Scott Baker, *Should We Pay Federal Circuit Judges More?*, 88 B.U. L. REV. 63 (2008); Stephen J. Choi et al., *Are Judges Overpaid?: A Skeptical Response to the Judicial Salary Debate*, 1 J. LEGAL ANALYSIS (forthcoming 2009), *available at* <http://ssrn.com/abstract=1077295>. On judicial appointment systems, see Stephen J. Choi et al., *Which States Have the Best (and Worst) High Courts?* 4 (Univ. of Chicago Law & Econ., Olin Working Paper No. 405, 2008), *available at* <http://ssrn.com/abstract=1130358>.

⁵⁰ See *supra* Part III(A).

⁵¹ See Choi et al., *Are Judges Overpaid?*, *supra* note 49, at 4; see also discussion *infra* Part III(B).

⁵² E.g., Ross Levine & David Renelt, *A Sensitivity Analysis of Cross-Country Growth Regressions*, 82 AM. ECON. REV. 942, 959 (1992); Howard Pack, *Endogenous Growth Theory: Intellectual Appeal and Empirical Shortcomings*, 8 J. ECON. PERSP. 55, 57 (1994).

judges.⁵³ Largely coming out of political science, the focus in this literature is primarily on the U.S. federal court system and does not attempt to compare court systems or individual judges in terms of their comparative levels of bias. Part of the reason for this, perhaps, is that while political bias is arguably the most salient measure of judicial independence in the U.S. federal system, other measures of independence such as corruption and fear of retaliation from the government are more important in other parts of the globe, and particularly so in the developing world. Given our focus on studies that attempt to measure and compare systems, therefore, we note this political bias literature only to the limited extent that it has been used in the comparative systems literature, which is in comparing the state systems in the U.S.

B. Electoral versus Non-Electoral Selection of Judges

A separate and smaller strain of empirical literature has been interested in the question of whether elections, appointments or merit selection systems produce the most effective court systems. The U.S. appears to be the only country that has enough variance in the types of systems used across different states such that this literature has been focused exclusively on the U.S.

While the literature on whether election, appointment or merit selection works best to select judges is large, much of it is not empirical. Initial attempts to empirically measure the performance of the different systems included cross-system comparisons of educational qualifications, prior legal experience, and even diversity levels.⁵⁴ The ready objection to these comparisons is that it is by no means clear whether any one of these variables translates into judicial quality.

A recent set of articles, confronting the measurement question rigorously, uses a variety of tacks to compare the state systems in terms of judicial independence levels.⁵⁵ The premise in most of these studies is that independence is the key aspect of a well-functioning judicial system. We describe some of the measures used by both political scientists and economists below.

Political scientists, seeking to compare elected versus appointed judges, have tapped into what one might think of as traditional assumptions regarding voter concerns about judges. Given the frequent rhetoric about “activist” judges

⁵³ E.g., Tracey George, *Developing a Positive Theory of Decision Making on U.S. Courts of Appeal*, 58 OHIO ST. L.J. 1635, 1640 (1998); Stephen J. Choi et al., *Professionals or Politicians: The Uncertain Empirical Case for an Elected Rather Than Appointed Judiciary* (Univ. of Chicago Law & Econ., Olin Working Paper No. 357, 2007), available at <http://ssrn.com/abstract=1008989>.

⁵⁴ Damon Cann, *Beyond Accountability and Independence: Judicial Selection and State Court Performance*, 90 JUDICATURE 226, 229-31 (2007); Bradley Canon, *The Impact of Formal Selection Processes on the Characteristics of Judges — Reconsidered*, 6 L. & SOC. REV. 579, 583-86 (1972); Henry Glick & Craig Emmert, *Selection Systems and Judicial Characteristics: The Recruitment of State Supreme Court Judges*, 70 JUDICATURE 228, 228-35 (1987).

⁵⁵ Cann, *supra* note 54; Aman L. McLeod, *Judicial Performance Review: A Balance Between Judicial Independence and Public Accountability*, 34 FORDHAM URBAN L.J. 315 (2007); Joanna Shepherd, *The Influence of Retention Politics on Judges' Voting*, 38 J. LEGAL STUD. (forthcoming 2009), available at <http://ssrn.com/abstract=997491>.

who are too lax on criminal defendants, one might expect that judges seeking the approval of voters will be tougher on crime than those who can rule without concern about voter approval. And in some studies, appointed judges were found more likely to favor criminal defendants than elected judges.⁵⁶ In the same vein, scholars also have found that the behavior of elected judges differs significantly from those of non-elected judges in the death penalty context.⁵⁷ Tapping into another traditional concern regarding elected judges—that they are beholden to financial contributors—scholars also have assessed whether judges in electoral states show favoritism toward business interests that have contributed toward their election campaigns.⁵⁸

Economists, seeking to get at the same question, have used different measures. One such measure has been to compare tort awards against out-of-state corporations in electoral states versus those in non-electoral ones. The former being higher, the argument is that the incentives of elected judges are to redistribute wealth from out-of-state corporations to in-state voters and to please the local trial bar.⁵⁹ Scholars also have compared litigation rates in electoral versus non-electoral states. They argue that the greater amount of litigation in non-electoral states can be explained by the fact that judges in election states are under more pressure to retain their jobs and therefore decide cases more consistently.⁶⁰ In a related study, the size of bureaucracies in electoral versus non-electoral states were compared.⁶¹ The larger size in non-electoral states, it is argued, is related to independence levels of the judiciary because government agencies in non-electoral states have to engage in more defensive efforts to protect themselves against judges who are unconstrained by the voters. Employment discrimination claims have been found to be more numerous in electoral than in non-electoral ones. Here, the explanation is that elected judges, seeking voters' approval and lacking independence from the voters, are more

⁵⁶ DANIEL R. PINELLO, THE IMPACT OF JUDICIAL-SELECTION METHOD ON STATE SUPREME COURT POLICY: INNOVATION, REACTION, AND ATROPHY 99 (1995).

⁵⁷ E.g., Paul Brace & Brent D. Boyea, *State Public Opinion, the Death Penalty, and the Practice of Electing Judges*, 52 AMER. J. POL. SCI. 360 (2008).

⁵⁸ See Damon Cann, *Justice for Sale? Campaign Contributions and Judicial Decision Making*, 7 ST. POL. & POL'Y Q. 297 (2007) (determining that campaign contributions are correlated with judges' decisions); Stephen J. Ware, *Money, Politics, and Judicial Decisions: A Case Study of Arbitration Law in Alabama*, 25 J.L. POL. 645 (1999). McCall and McCall also find an association between contributions and decisions but do not attempt to substantiate a causal relationship. See Madhavi M. McCall & Michael McCall, *Campaign Contributions, Judicial Decisions, and the Texas Supreme Court: Assessing the Appearance of Impropriety*, 90 JUDICATURE 214 (2007); see also Margaret S. Williams & Corey A. Ditslear, *Bidding for Justice: The Influence of Attorneys' Contributions on State Supreme Courts*, 28 JUSTICE SYS. J. 135 (2007).

⁵⁹ See Alexander Tabarrok & Eric Helland, *Court Politics: The Political Economy of Tort Awards*, 42 J.L. & ECON. 157, 186 (1999).

⁶⁰ F. Andrew Hanssen, *The Effect of Judicial Institutions on Uncertainty and the Rate of Litigation: The Election Versus Appointment of State Judges*, 28 J. LEGAL STUD. 205, 212 (1999).

⁶¹ F. Andrew Hanssen, *Independent Courts and Administrative Agencies: An Empirical Analysis of the States*, 16 J.L. ECON. & ORG. 534, 535-38 (2000).

likely to rule in favor of employees than appointed judges.⁶² Finally, scholars also have examined whether judges, whatever their party affiliation, adjust their behavior to please “retention agents”—those who decide whether a judge will be retained or not. The finding is that judges are more likely to vote in favor of traditional Republican interests when retention agents are Republicans, and are more likely to vote in favor of traditional Democratic interests when retention agents are Democrats. And the effect is larger when the retention process is electoral than when it involves reappointment.⁶³

A handful of articles have sought to go beyond just measuring independence and have looked at broader measures of judicial performance. The argument in favor of looking to broader measures is simple—that it is not enough to have judges be independent because independent judges who are incompetent or lack diligence will not provide quality judging. In this vein, although seeking to compare electoral and non-electoral systems was not their primary goal, Landes and Posner’s classic 1980 article found that citations (including federal and out-of-state citations) of state high court opinions were uncorrelated with selection system.⁶⁴ Also attempting to get at broader measures of judicial quality than aspects of independence, scholars have turned to survey measures. These have included surveys of senior attorneys at large companies evaluating the different states in terms of their attractiveness for business and surveys of judges rating their own state systems.⁶⁵ The surveys are attractive in that they are broad measures. But they have problems too because they are subjective evaluations by actors who not only have biases but who also do not always have the information necessary to perform a comparative analysis of the various systems.

Seeking to get at a broader measure for judicial quality, taking into account both insights related to judicial incentives in job retention and likely political biases driven by party affiliation, legal scholars Choi, Posner and Gulati use a composite measure that combines three different aspects of the judicial task: effort, quality and independence.⁶⁶ As proxies for the three aspects of the job they use the length and number of published opinions (effort), numbers of outside citations (quality) and frequency of open disagreement with judges of the same party affiliation (independence). Using this combined measure for effort, quality and independence, the various states are compared to see whether elected judges perform differently than non-elected ones. There do turn out to be differences in relative performance, but the strong suggestion of much of the

⁶² Timothy Besley & A. Abigail Payne, *Implementation of Anti-Discrimination Policy: Does Judicial Selection Matter*, CEPR Discussion Papers 5211 (2006), available at http://socserv.mcmaster.ca/payne/papers/besleypayne_May_2006_JLE_Revision.pdf.

⁶³ Shepherd, *supra* note 55.

⁶⁴ William M. Landes & Richard A. Posner, *Legal Change, Judicial Behavior, and the Diversity Jurisdiction*, 9 J. LEGAL STUD. 367 (1980).

⁶⁵ Daniel Berkowitz & Karen Clay, *The Effect of Judicial Independence on Courts: Evidence from the American States*, 35 J. LEGAL STUD. 399, 412-13 (2006); Cann, *supra* note 54, at 26.

⁶⁶ Choi et al., *supra* note 49, *Which States Have the Best (and Worst) High Courts?*, at 4.

scholarship that elected judges perform worse than non-elected ones is not borne out.⁶⁷

C. Ranking Judges and Courts

The two sets of measures discussed above were aimed at helping to answer particular theoretical questions: first, whether legal institution quality impacts economic growth and, second, whether elected judiciaries perform better than non-elected ones. As a function of the theoretical starting points of each inquiry, the measures took a particular shape. The third body of literature is primarily concerned with ranking and comparing individual judges. The researcher might, for example, be trying to determine measures that would reveal greatness in Supreme Court Justices. For the most part, given concerns with possible bias in the rankings, researchers have sought to use objective and verifiable measures—the dominant category being citation counts, which come in various forms.

The starting point for this scholarship is Mott's 1936 article.⁶⁸ Seeking to rank the various U.S. state high courts in terms of their relative quality, Professor Mott used three different measures. The measures included surveying law professors about what they thought of the different state courts, counting the numbers of cases from the different states that were included in casebooks, and comparing the number of inter-state citations obtained by each state high court.⁶⁹ The next important article in this line was written by Caldeira.⁷⁰ Also seeking to rank the state courts, Professor Caldeira used only the citation-count measure (albeit modifying it so that no state was allowed to disproportionately influence the measure).⁷¹ Looking to go beyond a simple ranking though, the focus of his inquiry was into the factors that result in high citations for some states and low ones for others.⁷² Later in other papers, he also sought to examine the factors determining why courts cited each other.⁷³ Recently, scholars have sought to update the Mott and Caldeira results and also to fine-tune the measures by, for

⁶⁷ Critics, including many judges, point out that these measures are too rough to properly evaluate the complex task of judging. See, e.g., Brannon P. Denning, *Empirical Measures of Judicial Performance: Thoughts on Choi and Gulati's Tournament of Judges*, 32 FLA. ST. U. L. REV. 1123, 1124 (2005); Daniel A. Farber, *Supreme Court Selection and Measures of Past Judicial Performance*, 32 FLA. ST. U. L. REV. 1175, 1176 (2005); Hon. Jay S. Bybee & Thomas J. Miles, *Judging the Tournament*, 32 FLA. ST. U. L. REV. 1055, 1056 (2005); Posting of Lawrence B. Solum, *A Tournament for Judges. Mad? Brilliant? Clever?*, Legal Theory Blog (Apr. 17, 2003), http://lsolum.blogspot.com/archives/2003_04_01_lsolum_archive.html#200162580.

⁶⁸ Rodney L. Mott, *Judicial Influence*, 30 AM. POL. SCI. REV. 295 (1936).

⁶⁹ *Id.* at 299.

⁷⁰ Gregory A. Caldeira, *In the Mirror of the Justices: Sources of Greatness on the Supreme Court*, 10 POL. BEHAV. 247 (1988).

⁷¹ *Id.*

⁷² *Id.*

⁷³ See, e.g., Gregory A. Caldeira & John R. Wright, *The Discuss List: Agenda Building in the Supreme Court*, 24 L. & SOC'Y REV. 807 (1990); Gregory A. Caldeira, *Commentary on Senate Confirmation of Supreme Court Justices: The Roles of Organized and Unorganized Interests*, 77 KY. L.J. 531 (1989).

example, using different types of citation counts.⁷⁴ Those have included broader measures that looked not only to citations by state courts to each other, but also citations by a variety of other courts and in treatises and law reviews. Among the narrower measures used have been the Shepherd's designation of cases being "followed," in effect for a citation being especially strong.⁷⁵

Of a more recent vintage is the line of scholarship seeking to rank individual judge performance. This scholarship has largely focused on the U.S. federal courts. Caldeira's work once again was pioneering. In 1988, he examined the question of how to quantify judicial greatness.⁷⁶ To determine greatness (and the factors that produced it), he tallied the names mentioned in various non-empirical scholarly discussions of the great justices.⁷⁷ Also attempting to quantify greatness, Posner, in an analysis of Benjamin Cardozo used a variety of citation measures to quantify Justice Cardozo's reputation.⁷⁸ Posner later applied a similar methodology in his analysis of Learned Hand.⁷⁹ Following in this vein have been multiple studies that compared the performances of U.S. federal courts of appeals judges.⁸⁰ Here also, scholars have gone beyond the basic citation counts, by eliminating negative citations, string citations, and boilerplate citations, and by looking at measures such as invocations (the number of times a judge is explicitly mentioned by name, in the context of a citation).⁸¹ Similar methodologies also have been applied to analyze judicial performances in Australia and New Zealand.⁸² Unpacking citation counts further, Choi and Gulati have sought to measure the degree of bias in citations (such as whether judges engage in reciprocal citations or biased citations based on political affiliation). Those bias estimations for individual judges were then used to construct a ranking of the judges in terms of the different levels of citation bias.⁸³

⁷⁴ Jake Dear & Edward W. Jessen, 'Followed Rates' and Leading State Cases, 1940–2005, 41 U.C. DAVIS L. REV. 683, 686–88 (2007); Choi et al., *Which States Have the Best (and Worst) High Courts?*, *supra* note 49, at 4.

⁷⁵ Dear & Jessen, *supra* note 74, at 683.

⁷⁶ Caldeira, *supra* note 70, at 247–48.

⁷⁷ *Id.* at 250.

⁷⁸ RICHARD POSNER, CARDOZO: A STUDY IN REPUTATION 74–91 (1990).

⁷⁹ Richard Posner, *The Learned Hand Biography and the Question of Judicial Greatness*, 104 YALE L.J. 511 (1994) (reviewing GERALD GUNTHER, *LEARNED HAND: THE MAN AND THE JUDGE* (1994)). See also Lee Epstein et al., *Rating the Justices: Lessons from Another Court* (Apr. 1992), <http://epstein.law.northwestern.edu/research/conferencepapers.1992MPSA.pdf> (using basketball analogies to quantify comparative greatness among Justices).

⁸⁰ See, e.g., Baker, *supra* note 49.

⁸¹ David Klein & Darby Morrisroe, *The Prestige and Influence of Individual Judges on the U.S. Court of Appeals*, 28 J. LEGAL STUD. 371, 375–76 (1999).

⁸² See, e.g., Dietrich Fausten et al., *A Century of Citation Practice on the Supreme Court of Victoria*, 31 MELB. U. L. REV. 733, 744–45 (2007); Paresh Narayan & Russell Smyth, *What Explains Dissent on the High Court of Australia? An Empirical Assessment Using a Cointegration and Error Correction Approach*, 4 J. EMPIRICAL LEGAL STUD. 401 (2007).

⁸³ Stephen J. Choi & G. Mitu Gulati, *Ranking Judges According to Citation Bias (As a Means to Reduce Bias)*, 82 NOTRE DAME L. REV. 1279, 1303–06 (2007).

Seeking to broaden the evaluation of judicial performance beyond citations, Choi and Gulati combined measures for effort (publication rates), quality (citations) and independence (disagreement with same party judges as compared to disagreement across party lines).⁸⁴ Other scholars, seeking to improve the Choi and Gulati measures, added in measures such as rates of reversal of the various appeals court judges by the Supreme Court.⁸⁵

Cross-country comparisons in this line of scholarship are rare. That may be because the favored measures of those in the ranking business—citations—depend crucially on judges issuing opinions where the authors are identified. Many countries, however, are not in the business of judges issuing individual opinions, let alone dissents. To rank courts across countries, therefore, other measures need to be devised. Again, Caldeira's research is at the forefront of cross-country comparisons of judicial/court value. Along with Gibson and Baird, Caldeira used surveys of members of the general public to determine a court's level of institutional support from the public.⁸⁶ The goal in this research is to determine those factors that lead to public views about court legitimacy.⁸⁷

We have by no means exhausted the list of measures used. For example, in seeking to determine whether higher judicial salaries translate into better judicial performance, Baker estimated comparative measures for both the time it took various U.S. federal appeals court judges to decide a case and also their relative degree of bias when voting on politically charged cases.⁸⁸ Elsewhere, scholars have compared the voting patterns of judges appointed by different patrons, for example, to determine whether judges appointed by, for example, President Clinton, were more liberal than those appointed by President Reagan or either President George H. Bush or George W. Bush.⁸⁹

The various measures used are all flawed—each capturing a narrow perspective of the full portrait. But it is difficult to claim that these various studies, when put together, do not shed light on judicial performance and institutional design. One can quibble with the conclusions, but common law jurisdictions appear to do justice differently from civil law ones and elected judges respond differently to incentives than their non-elected counterparts. And while one can argue about whether citation counts, invocations or the number of cases that enter the casebooks measure anything close to quality (or instead merely measure vanity or judicial self-marketing), the fact remains that when one judge has five times the number of citations than her colleagues, one can conclude that the judge is likely writing opinions that are more lucid, more adventuresome, more creative, more scholarly, more well-reasoned or more

⁸⁴ But cf. Choi et al., *Are Judges Overpaid?*, *supra* note 49, at 20-23.

⁸⁵ Frank B. Cross & Stefanie Lindquist, *Judging the Judges*, 16-17 (unpublished manuscript on file with authors).

⁸⁶ James L. Gibson et al., *On the Legitimacy of National High Courts*, 92 AM. POL. SCI. REV. 343, 348-55 (1998); James L. Gibson & Gregory A. Caldeira, *Changes in the Legitimacy of the European Court of Justice: A Post-Maastricht Analysis*, 28 BRIT. J. POL. SCI. 63, 74-85 (1998).

⁸⁷ *Id.*

⁸⁸ Baker, *supra* note 49, at 86-89, 101-105.

⁸⁹ Glick & Emmert, *supra* note 54.

readable than her compatriots. Not all of these attributes may be good, but something is happening which in turn can shed light on how the institution is working.

IV. ALTERNATE MEASURES OF JUDICIAL PERFORMANCE: THE FEDERAL TRIAL COURTS

We began Part II with a set of basic institutional design questions about courts. As discussed in Part III, the academic literature on judicial measurement has attempted to get at a handful of these questions such as whether elected judges perform better than appointed ones. But, overall, the surface has barely been scratched in terms of our knowledge about the optimal institutional design of the judiciary. In this section, we propose a handful of new measures aimed explicitly at the kinds of institutional design questions that we think would interest judges and focused in significant part on the United States district courts.⁹⁰

We choose to focus primarily on the United States district courts for three reasons. First, it is the trial courts where most law happens. That is, the trial courts are the final arbiters of the overwhelming majority of disputes. Second, given the relative lack of scrutiny that the lower courts receive, and the greater numbers of judges and districts as compared to the circuit courts, this is also where differences in institutions are most likely to show up. Third, the literature on judicial measurement has largely ignored the question of how best to measure trial court performance.

More concretely, imagine a researcher seeking to compare the performance of the Indian judicial system to that of the United States or United Kingdom. The easiest tack to take might be to compare the Supreme Courts of the relevant jurisdictions—these are certainly the courts about which the best information is likely to be available. We suspect that although there would be differences, the performances of the justices of the Indian Supreme Court would be similar to their American or English counterparts. But we predict that the differences in the performances of the trial courts in these countries will be much greater—precisely because it is in this location that differences in court resources, judicial selection mechanisms, and caseload, to name a few, are likely to show up. If researchers are, therefore, interested in whether the common law courts of England and the United States are working better for business groups than the civil law courts of Germany and France, it is the trial courts that are most relevant, not the high courts where business disputes are rarely heard. A similar argument applies to inquiries into whether elected judiciaries work better than non-elected ones—the difference in sheer number of judges and actual dispositions available for study would suggest that a more complete picture would result from a comparative analysis of trial courts. This is not to say that

⁹⁰ In Part IV(F) *infra*, we offer some measures specific to assessing a judge's independence from prosecutors that might be more useful in the state court context and in Part IV(H) *infra*, we offer some measures specifically suited to appellate courts.

comparing the performances of the high courts is irrelevant, but simply that a comparison of the performances of the trial courts is likely to be more revealing and lead to more significant knowledge about institutional design.

In proposing these measures for the federal trial courts, we attempt to take the perspective of a judge who is interested in how the data might help shed light on institutional reform. In our experience, most judges are interested to learn how other judges handle their cases, organize their chambers, and manage their courts. Because judges, for the most part, do not compete with one another for promotions or salary increases, they do not feel the same constraints that other professionals might feel in discussing the virtues of their internal procedures. And most judges that we know are keenly interested in whether there are better ways to organize themselves which might lead to greater efficiency or fairness. Whenever district judges gather there will be discussion of such topics as: how best to pick a jury, how to deal with the complexities of class actions, whether to hire a secretary or an additional law clerk, whether to hire a career law clerk or a term clerk, how best to use magistrate judges, and how to sentence during this period of uncertainty under the sentencing guidelines. Appellate judges have similar discussions.

Judges should be interested in having the topics of such intense informal discussion systematically studied so that firmer conclusions can be drawn about the efficacy of different approaches and techniques. Once it is established that certain practices are generally preferable, many judges will want to know how they compare against these best practices and will make individual decisions about whether to adopt new practices. Where the empirical evidence is clear, it may be expected over time that institutional pressure will be brought to bear on judges who follow the less effective practice, perhaps through local court or national Judicial Conference policies and rulemaking.

How might we identify those measures that could be equated with a best practice? This will require empirical study as to some of the measures. Before proposing new measures though, we briefly examine in Section A the applicability to the trial court of two frequently used measures for the appeals courts: publication rates and citation rates. In Sections B through F, we propose several ways of measuring judicial behavior at the trial court level. In Section G, we discuss new ways of measuring appellate judge behavior.

A. Publication and Citation Rates

A key element of the task of an appeals court judge is to explicate the law; that is, to explain her decisions in a manner that enables others to understand the rationale for the decision and to be guided or bound by that explanation in future cases or circumstances. Publication of reasoned decisions, therefore, is a key element of the job. Judges who are publishing many fewer or shorter decisions than others are doing something differently. For much the same reasons, citation rates also are arguably relevant for assessment of an appeals court judge's performance. If part of the appellate judge's task is to explain the law in published decisions so that others can use and be bound by those explanations, then a comparison of how much the decisions of the different judges are actually

being used should shed light on whether that aspect of the job is being performed fully.

For trial courts though, and certainly for the United States district courts, while providing written explanations of decisions is an important part of the job, doing so in “published” form is arguably not as significant an aspect of the job. As an initial matter, the decisions of a district judge are not binding on any other judge, not even the issuing judge presiding over a different matter. Moreover, it is not even clear what it means to “publish” a decision in the district court. Now that all opinions must be made available on-line on the courts’ websites,⁹¹ and given that district court opinions have no precedential value, one can ask: What is the point of a district judge designating an opinion for publication in the Federal Supplement or one of the on-line services? From a measurement perspective, the answer is that a trial judge choosing to send her opinion to Westlaw for publication has made a choice largely independent of the requirements of the job. The follow-up question then is: what does this choice signify? Perhaps by designating the opinion for “publication” the district judge vouches for the opinion—signaling confidence in the opinion and its usefulness to other judges—and, therefore, increases the likelihood that others will come across it.⁹² Most likely, the judge in question has worked harder on polishing and editing the opinion than she would have had she not been planning to designate it as “for publication.” One can infer from the foregoing that, at least along one axis—effort put into polishing and editing the statement of reasons in a case—and adjusting for caseloads, the judges choosing to publish more opinions are working harder on their opinions. And this is not an unimportant axis. Litigants probably value a better articulated statement of reasons in the court’s decision. It not only reassures the litigants that the judge thought hard about their case, but the clarity of rationale for the decision makes errors more obvious and enables an easier appeal than if the litigants had to speculate about the trial judge’s rationale. There are also general public good values to having better articulated statements of the law—namely, that lawyers and other judges are provided with guidance. Finally, although this may be a stretch, it is possible to infer that the judges who exert more effort along the publication axis are also likely to be exerting more effort along other axes. That is, they are all-around hard workers.

But there are other possibilities. Some judges may enjoy having their opinions published by the leading law publishers, regardless of quality. That is, the publication designation might not be an indicator of higher quality and, therefore, higher usefulness to other judges, but rather simple self-promotion. Alternatively, and more interesting, it may be that judges who publish a lot are

⁹¹ E-Government Act of 2002, Pub. L. No. 107-347, 116 Stat. 2899 (codified as amended in several sections of 44 U.S.C. (2005)).

⁹² Alternatively, it may be that some trial judges are auditioning for promotion to the appeals court and wish to demonstrate skill at the type of task that an appeals court judge is expected to perform regularly. An interesting question to ask using publication and citation rates, therefore, might be whether the trial judges who publish and are cited the most are indeed the ones who get promoted to the appeals court. This seems unlikely, particularly since in the context of judicial appointments, a written record sometimes hurts a candidate for promotion or appointment.

spending more effort on polishing their opinions and less effort on other judicial work and values, such as getting the decision out in timely manner, presiding over other trials and deciding other motions. It would be fairly simple to measure whether published opinions take more time than unpublished.⁹³ If we could also devise a means to measure effort spent on trials and motions, which we might do using various measures of efficiency, then the question of whether high publication rates result in less effort on trials and motions could be examined. If it does turn out to be the case that high rates of publication are invariably associated with lower effort on trials and motions, and with a significantly longer time to decision, that would seem to suggest a problem with the decision to publish. And from an institutional design perspective, it might suggest that judges should not be given as much unguided discretion in terms of choosing what to publish.

The above discussion assumes that the judges under consideration have comparable case loads. When the judges do not have comparable caseloads, differences in publication rates might be an indicator of a variety of things.⁹⁴ For judges with lower or less complex caseloads, the decision to fill out their day by polishing up their written opinions might show their industry.⁹⁵ But for judges in jurisdictions with higher or more complex caseloads and larger backlogs of cases, the expenditure of resources on publication might be wasteful from an institutional design perspective. For district judges with similar medium-to-high caseloads, we might look to citation rates not simply as a measure of the judge's opinion-writing ability, but also as a way to evaluate the judge's decision to spend time polishing and publicizing a decision rather than performing other judicial tasks. So a high designation for publication rate, particularly when coupled with a low citation rate, might suggest that the judge is not spending her time in the optimal fashion. And a higher-than-normal reversal rate for a judge's published decisions might indicate a particularly poor use of time. On the other hand, published decisions may address unsettled areas of the law and hence would be more likely to result in disagreement and reversal.⁹⁶

Publication rates can also yield information in terms of strategic behavior. It is possible, for example, that a judge may want to influence an area of the law of particular interest to the judge. If so, such a judge may focus his publication

⁹³ While it would be time-consuming and impractical to track the precise number of hours spent drafting an opinion, the time to decision might serve as an appropriate measure of this.

⁹⁴ See Ahmed Taha, *Publish or Perish? Evidence of How Judges Allocate Their Time*, 6 AMER. L. & ECON. REV. 1, 1 (2003) (finding that workload differences correlate with publication differences).

⁹⁵ See generally *id.* at 20. Alternatively, judges in low caseload districts may make themselves available to take cases from districts that have higher caseloads. This is easily accomplished within the same circuit. Any evaluation of the comparative benefits of how judges spend their time should take into account whether the judge could be assigned additional cases from other districts.

⁹⁶ It is possible that by publishing an opinion the district court is signaling the importance of the case to the appellate court as well as creating the possibility that other district judges will rely on the opinion. An additional measure of study could be whether a district judge's consideration of publication should include a calculation of whether the very fact of publication may increase the likelihood of reversal.

efforts in specific areas. Or, to the extent publication increases the possibility of reversal, judges might choose to publish more or less depending on the identities of those on the relevant court of appeals.⁹⁷

Judges, we suspect, often do not have a clear sense of whether they should be publishing more or less and how important publication is to the performance of their job. They also probably do not have a sense of whether their colleagues are using the choice to publish in a strategic fashion (and we do not mean to use the term “strategic” pejoratively). Information about the differences in practices would be useful in generating a debate among judges and policy makers about how best to improve the system.

B. The Setting of Trial Dates

In its study of the Civil Justice Reform Act and the various pilot programs that were established under the Act, the 1997 RAND report found “that an early and firm trial schedule, combined with limited time for discovery, can reduce delay in complex civil litigation without increasing costs.”⁹⁸ This finding was not controversial at the time and continues to ring true. A judge intent on improving her handling of civil litigation might want to know whether by comparison to other judges she tends to set a firm trial date at an early point in the litigation or whether, again by comparison, she is more prone to grant continuances or to re-set the trial date because of the demands of her own caseload. She may wish to know whether she is setting the same kinds of discovery time limits as other judges in comparable cases. Given the importance of these two management tools, judges might wish to know whether they or members of their staffs, such as calendar clerks, are effectively managing their civil dockets, at least by comparison to what other judges with similar caseloads are doing.

Based on the foregoing, one measure of judicial efficiency at the trial level might be to answer whether a judge or court sets an early and firm trial date with a limited time for discovery. Studies, like the RAND study, might further refine this measure so that we could determine optimal times for discovery and for pre-trial activity depending on the type of case. Even until such refinement, judges would be eager to know whether they are out of step with other judges in comparable courts or with other judges in their same court. Such a measure could be used to compare courts and judges across jurisdictions. It could also be used to measure different judicial selection methods: for example, it is possible that elected judges are less independent of the bar and therefore more willing to

⁹⁷ James Brudney & Deborah Jones Merritt, *The Influence of Appellate Judges' Social Backgrounds When Reviewing NLRB Decisions*, 2 EMPLOYEE RIGHTS Q. 13 (2002) (labor law cases); David S. Law, *Strategic Judicial Lawmaking: Ideology, Publication, and Asylum Law in the Ninth Circuit*, 73 U. CIN. L. REV. 817, 819-20 (2005) (discussing the possibility of strategic choices at the trial court levels); Russell Smyth, *Judges and Academic Scholarship: An Empirical Study of the Academic Publication Patterns of Federal Court and High Court Judges*, 2 QUEENSLAND U. TECH. L. & JUSTICE J. 198 (2002).

⁹⁸ RAND, THE CIVIL JUSTICE REFORM ACT OF 1990: FINAL REPORT 3 (1997).

grant continuances at the request of any lawyer. Because lawyers generally do not bear the main cost of delay—the parties usually do—and because motions for continuance are often the result of some conflict that the lawyer has, the judge who insists on a firm trial date will make few friends in the bar and probably gain little compensating support from the litigants who tend to receive information about their case through the lens of their lawyer.

C. The Use of Law Clerks

With additional study, similar case management or chambers management techniques might be identified and raised to the level of a best practice and hence a measure of judicial excellence. We might study how judges use their law clerks; whether they use the clerks primarily to do background research or whether they use them in the drafting of opinions. We might try to evaluate these different approaches. Or we might want to know whether law clerks with actual legal experience perform better than those straight out of law school. Those judges who hire career law clerks believe that by avoiding the yearly or bi-yearly turnover of clerks, they save significant amounts of time in selecting and training their clerks and that more experienced law clerks are more efficient and exercise better judgment. Judges who prefer term clerks often concede the efficiency of the career clerk; indeed, they fear this efficiency and the possible reliance it could bring. A persuasive study of these issues relating to law clerks might require surveys, interviews, and perhaps internal observational studies of the workings of chambers and the decisional process. There are other similar kinds of internal organization that would bear study and evaluation, such as the use of magistrate judges or master calendaring systems:

D. Speed of Decision

We would probably agree that a good trial judge decides pending motions expeditiously, even though we could not fix an exact optimal time for all motions for all similarly-situated judges. Here, we would be content to know whether individual judges are resolving such motions consistently with colleagues who have similar weighted caseloads. The judge who is either significantly faster or significantly slower might be considered either insufficiently deliberative or insufficiently decisive and efficient. In an effort to provide oversight of the speed to decision, federal judges now must report their “six month lists”—the number of motions in civil cases that have been under submission for over six months since the filing of the motion.⁹⁹ The “six month list” is almost useless: the time period is triggered by filing, not by when the motion is fully briefed and ready for argument or decision. But giving judges an assessment of whether they are resolving submitted motions ready for decision as expeditiously as other judges managing similar caseloads would be useful information for judges. A judge too far off the median might reconsider how he does the job, whether he

⁹⁹ See 28 U.S.C. § 476 (2005).

needs to work harder or faster. Similarly, a comparison as between different courts of the average or median time to decision of similar motions could lead to some conclusions about different systems and their design. For example, if courts without adequate staffing, such as law clerks, take significantly longer to resolve submitted motions, then we have learned something about institutional design. Accordingly, another useful measure for trial judges and courts might be an assessment of the average and median time to decision of completed motions and bench trials.

E. Sentencing

Sentencing is another area in which the average or median sentence and the degree of variation have inherent normative stature since most would agree that one aspect of a good sentencing system provides for a certain degree of uniformity as between judges who sentence similarly-situated defendants for the same crimes. In the federal system, prior to the Sentencing Reform Act, it was common for pre-sentence reports, prepared by probation officers, to provide the judge with median sentences, and percentiles, for the crime at issue both by nation and by judicial district. Judges found the information useful. Thereafter the sentencing guidelines served some of the same function; however, with the new approach to sentencing under *Booker*,¹⁰⁰ and the corresponding increase in the frequency of departure from the guidelines, it may be that judges would find the median and percentile information helpful once again. Interestingly, sentencing data is not provided on a single judge's sentencing history, although the information could easily be compiled by the United States Sentencing Commission. Such information could be useful to the individual judge who would want to know whether he has been out of step with colleagues in other courtrooms, particularly those in the same courthouse. The judge might also want to know whether he has been internally consistent, because such consistency is an aspect of justice and helps guard against improper or inadequately-considered factors. A judge might particularly want to know under what circumstances judges on the same court are willing to deviate from the norm, and she might wish to know her own history on departure so as to attain at least a greater degree of internal consistency when deciding to depart from the norm. The absence of such information probably reflects political reality; judges would like the information but not if it also would be provided to members of Congress or members of the bar. Perhaps some members of either group would use the information for improper purposes. Nonetheless, on balance, we think that the information is too valuable not to collect and make available at least to the judge, if not to others.

¹⁰⁰ 543 U.S. 220 (2005).

F. Independence

Another normative measure important in the trial court is the judge's independence from the parties and the bar. There are different ways to get at this, but we would all agree that a judge should not feel coercive pressure from any litigant or member of the bar. This is particularly of concern in criminal cases. We could measure the degree of authority that the prosecutor has over the judge. For example, we could look at whether the prosecutor can peremptorily challenge the judge, effectively removing the judge from all criminal cases. Such power exists in California and is used by prosecutors to "school" judges who are deemed unfriendly to the prosecutors' office.¹⁰¹ Or we might look at whether the prosecutor chooses the judge before whom a case will be tried and also decides when the case will go to trial. The recent Duke Lacrosse case is an example of how these powers can be abused.¹⁰² Finally, we might look at whether judges who are considered too demanding of the prosecutor or too favorable to the defense are routinely challenged at elections by prosecutors who wish to become judges and whether such challenges have the de facto support of the district attorney. Therefore, we propose that a possible measure of the independence of the judiciary could be constructed according to whether the judge can be removed by peremptory challenge by the prosecutor, whether the prosecutor indirectly or directly can choose the judge to hear specific cases, whether the prosecutor determines the timing of the trial, and whether the prosecution office will field a candidate to challenge a judge who is viewed as too lenient or defense oriented. From these measures a composite score for judicial independence in criminal cases could be assembled.

G. Appeal and Reversal Rates

We might consider whether appeal rates tell us much about the quality of the trial court. If some judges are having their decisions appealed at an exceptionally high rate, as compared to other judges in the same jurisdiction, that could suggest that those are poorly reasoned decisions producing erroneous outcomes. Again, there are other possibilities that would have to be examined as well. High appeal rates might also suggest that the judge writes cogent and well explained opinions that are easier to appeal. The high appeal rate could also be that judges on the appellate court appear to lack respect for the trial judge or disagree with his or her judicial philosophy (and that litigants, recognizing the conflict, appeal at higher rates).

Appeal rates can be parsed further for information by examining reversals as a function of the number of appeals. Reversals, at first cut, should be a finer measure of a judge's error rate. That said, to compare judges across the same

¹⁰¹ CAL. CIV. PROC. CODE § 170.6 (West 2004).

¹⁰² See *N.C. State Bar v. Nifong*, 06 DHC 35 (July 31, 2007) (Disciplinary Hearing Commission of the N.C. State Bar's Amended Findings of Fact, Conclusions of Law, and Order of Discipline), available at <http://www.ncbar.com/orders/06dhc35.pdf>.

appellate jurisdiction in terms of their reversal rates, one would have to correct for other factors potentially influencing these rates such as caseloads, trial rates, and the subject matter of the cases.

Despite these qualms, there is probably some optimum rate of reversal in any judicial system. We would not want the rate to be zero because that could mean that the appellate judges are not diligently reviewing for error and it seems unlikely that all lower court decisions are correct. But we would not like to see a reversal rate that is significantly higher than the norm because this suggests that either the appellate court or the lower court or both are not doing their jobs as well as other courts, thereby inflicting cost, delay, and uncertainty on the litigants.

The point is that while there are perils to using reversal rates, there is information to be gleaned from them. Judges with reversal rates above the norm are likely doing something different than those below the norm. Assuming that the more frequently reversed judges know how to write opinions so as to avoid reversal, the fact that they are being reversed more often might suggest a greater willingness to be reversed due to disagreement with the appellate court's approach to certain legal issues or poorer understanding of the direction the appellate panel is likely to take. Of course, a higher reversal rate might just demonstrate a less careful or intelligent district judge. To glean additional insight and separate the various possibilities, the data can be parsed further. One can examine, for example, whether a judge's reversal rate depends on whether the appellate panel is made of judges appointed by a president of the same political party or of the same gender or race. One could compare, for example, a judge's reversal rates by panels of appeals court judges who share her political affiliation or judicial philosophy to reversal rates of panels that are of a different affiliation or philosophy. And that data can be fine-tuned further by separating reversals where there is a split vote (that is, one judge above dissents) to those where there is a unanimous vote. The difficult or close cases presumably will be those where a split vote is more likely. In addition, the number of reversals where oral argument is granted can be compared to the number where it is not. The grant of oral argument is usually a sign that the court considers the issues raised by the case non trivial—although the norms differ according to circuit and the data would have to be corrected for that.¹⁰³ A judge whose reversal rate is different depending on the make up of the appellate panel is likely doing something different than a judge who has the same reversal rate regardless of the make up of those above.¹⁰⁴ Similarly, a judge whose reversals are predominantly by split panels is likely doing something different than a judge whose reversals are by uniform panels (particularly if those uniform panels are made up of judges of the

¹⁰³ A similar measure could also be used, perhaps more effectively, on the Supreme Court.

¹⁰⁴ Reversal rates can also be calculated as a function of specific subject areas because some subject areas are more politically-heated than others. If a judge has a high reversal rate in both the politically-heated areas and in the mundane ones, then perhaps an inference can be drawn that it is not politics or judicial philosophy that is at work. Instead, the reversals of this judge's decisions are exercises in error correction. If, however, the judge is being reversed only in politically-heated areas, then it is not a stretch to infer that politics or judicial philosophy is playing a role.

same judicial philosophy as the trial judge). The point again is that there is an enormous amount of information available to be gleaned from such measures.

H. The Appellate Courts

Turning to the appellate courts, as with trial judges, a useful measure of the efficacy of the court is time to decision from the point of full briefing. Although readily available appellate statistics would not permit comparisons between individual judges—one suspects, however, that appellate courts maintain their own internal data on how quickly each judge circulates opinions and either joins or otherwise responds to the proposed majority opinion—aggregate information would permit comparisons to other courts. Again, faster decision-making would not necessarily lead us to conclude that the judges were more efficient or diligent, but we would be interested in deviations from the norm for similarly-situated courts. We could study the effects of workload and staffing on the time to decision. With court cooperation, we could also study the effects of certain internal procedures on time to decision. For example, it is no secret that the California appellate courts prepare their opinions before the case is set for oral argument. They do this because by state law the judges of the court will not be paid unless the opinion is released within a set period of time following the argument.¹⁰⁵ Whether this leads to prompter decisions or simply delays the setting of the oral argument is open to study. Further, with court cooperation one might study the effect of different decision-making processes on outcomes. For example, if the California appellate courts prepare their opinions prior to the oral argument does this mean that the judges are less open to the substance of the oral argument and, if so, does this affect the quality of the opinion? Similarly, in both the state and federal court context, if only one judge's staff prepares a memorandum on a pending appeal, does this tend to encourage less scrutiny by the two other judges such that the benefits of sitting in a panel are diminished?

With court cooperation, we could look more critically at the intersection of opinion preparation and opinion quality. If judges would identify those opinions in which the primary drafter was the judge, rather than a clerk, we could look to see if there are differences in the citation rates for those opinions as opposed to those principally prepared by law clerks. And, as to those opinions principally prepared by law clerks, we might look to see if there are differences in the citation or reversal rates for opinions prepared by career law clerks as compared to those prepared by term clerks. Perhaps on some anonymous basis judges might permit the veil to slip if only because judges would be interested in the outcomes of such a study.

¹⁰⁵ CAL. CONST. art. VI, § 19 (permitting the withholding of a judicial salary when any cause "remains pending and undetermined for 90 days after it has been submitted for decision").

V. CONCLUSION

We have tried to suggest possible measures that are both practical and that would be of interest to judges and researchers alike. These measures can speak to institutional design and would give judges and others a basis on which to develop best practices. Our point is that there is much common ground on which researchers and judges may speak to one another. Certainly many who undertake empirical research wish not just to describe and evaluate the courts, but to provide information that is useful to judges and courts. Surely judges would welcome data that might confirm the value of current practices or suggest the benefits of reform. Let us begin.