Pitfalls of Measuring the Rule of Law

Tom Ginsburg, University of Chicago

Available at: https://works.bepress.com/tom_ginsburg/33/
Pitfalls of Measuring the Rule of Law

Tom Ginsburg*

The recent demand for new measures of the rule of law confronts several methodological challenges. This article calls for careful attention to fundamental social science ideas of conceptualization and measurement in approaching the rule of law. Efforts to measure complex social phenomena such as the rule of law are challenging, and thus require that researchers and policy makers pay attention to the cautionary rules of social science in their efforts. Violating these basic rules risks producing measures that are not reliable or valid, and could be a bad basis for policymaking. This paper demonstrates some of the pitfalls that rule of law researchers have fallen into and suggests improvements in measurement approaches.

Introduction

We live in an era of measurement and quantification, in which both the supply and demand of numerical indicators is greatly expanding. We rank cities for their level of global integration, universities for their level of quality, businesses for their environmental practices, and countries on a wide range of development measures. Naturally, the multibillion dollar effort to promote the rule of law has been part of this trend, as scholars and policymakers have sought to generate and utilize relatively objective measures of legal phenomena. We now have a host of different indicators – some say more than 150-purporting to capture ‘governance’ and related aspects of institutional quality, including the rule of law. The veritable thicket of measures – and the fact that they are not always tightly correlated-

* Leo Spitz Professor of International Law and Professor of Political Science at the University of Chicago School of Law.
poses challenges to social scientists and to policy makers who would like to use them to make comparative judgments about the quality of different legal systems and governance structures.

In this article, I identify several pitfalls in measuring the rule of law. First, I call for paying attention to fundamental social science ideas of conceptualization and measurement in approaching the rule of law. Efforts to measure complex social phenomena, such as the rule of law, are challenging, and thus require researchers and policy makers to pay attention to the cautionary rules of social science in their efforts. Violating these basic rules risks producing measures that are not reliable or valid, and could be a bad basis for policymaking. Too much is at stake to pretend that these challenges do not exist. First, I focus on two basic challenges: conceptualization and measurement. I call for paying attention to fundamental social science ideas of conceptualization and measurement in approaching the rule of law. In each area, I demonstrate some of the pitfalls into which rule of law researchers have fallen.

Next, I examine the problem of feedback effects, namely the problem that reporting measures under some circumstance might undermine the very efforts to improve the underlying phenomenon. This is likely to be the case in situations where perceptions are particularly important. Perceptions can produce their own reality because people’s strategies in social situations are influenced by the strategies they think that other actors are likely to play. This is a real risk for a phenomenon like the rule of law, though I illustrate it using the concept of corruption.

After the discussion of several potential pitfalls, I conclude on an optimistic note. I am not arguing that we should abandon the task of attempting to measure legal institutions. There are good reasons to want to measure legal phenomena, and the large-n statistical work that measurement facilitates has produced important insights. But the social science cautions must be heeded to prevent bad policy advice.

Conceptualizing What Is Being Measured

The first challenge derives from what is called conceptualization. This requires formulating a concept at a particular level of abstraction.\(^6\) A good social science concept has several features, including coherence, parsimony, and utility. It should be relatively clear and bounded from neighboring concepts; it should be rela-

---

tively parsimonious; susceptible in principle to measurement; and capable of being deployed within research designs.7

It is tempting to think that the more encompassing the concept, the more difficult the measurement challenge will be. However, this is not always the case. Democracy is a high level concept that nevertheless has been reduced, by some scholars, to an essential core of regular and contested elections, or alternation in power.8 The latter is fairly easy to measure. Others view the concept in more substantive terms, encompassing the rule of law, fundamental rights and minority protections, and substantive equality. Needless to say, adding these elements to the definition of democracy exacerbates the measurement challenge. To compare the quality of democracies under the expanded definition, one needs measures of each of the sub-components, and a mode of aggregating them into a single feature. Weighting various sub-components can be a source of bias in the measures. But these challenges have not prevented the emergence of various measures of democracy, as well as aggregate meta-measures.9

Efforts to measure the rule of law have been subject to criticism for poor conceptualization. This is particularly the case with the World Bank’s World Governance Indicators and its rule of law index. One critique is that it aggregates too many discrete elements into a single overarching concept. At one point, the World Bank defined the rule of law as ‘the extent to which agents have confidence in and abide by the rules of society, and in particular the quality of contract enforcement, the police and the courts, as well as the likelihood of crime and violence.’10 This definition conflates many different notions into a single concept, including both crime and contract enforcement in the same framework. The applied concept contains both procedural elements, consistent with Fuller’s widely accepted definition of the rule of law, with substantive concepts, such as security of the person and freedom from crime. Some of the elements may be within the control of government officials; others may have to do with underlying economic and social conditions unrelated to the rule of law. These things may not necessarily go together, introducing not only measurement error, but also conceptual confusion.

Another problem at the level of conceptualization is whether or not the rule of law concept is distinct from neighboring concepts. The World Governance Indicators for any given year are highly correlated; indeed, one estimate shows a .95 correlation between the rule of law and the separate indicator for government

---

7 Gerring, Social Science Methodology, above n. 6, p. 40.
effectiveness and a .91 correlation with regulatory quality.\(^\text{11}\) As Professor Thomas puts it, ‘the correlations are so high that they raise the question of whether the indicators are measuring distinct constructs.’\(^\text{12}\)

One of the reasons that measuring the rule of law is difficult is that the underlying concept is complex. It is not simply a matter of measuring, or even aggregating, the presence or absence of institutional structures. Rather, the rule of law is best thought of as a configuration of structures, social attitudes, and even traditions. This raises problems of aggregation, addressed in other contributions to this special issue. And here it is that measurement concerns arise. If the sources used to measure the various components of the rule of law are similar, aggregating them raises the possibility of correlated error, which would produce biased results. Indeed, this precise critique has been directed at the World Governance Indicators.\(^\text{13}\)

Furthermore, there is no single ideal formula to achieve it. It may be that, in some countries, an independent judiciary is a crucial element; in other countries the judiciary can become too autonomous and can itself become a major political actor. In some countries, prosecutors will be key actors for ensuring that the rule of law is upheld; in others, civil society might be more important. It is partly for this reason that efforts to simply transpose institutional structures have produced generally disappointing results.\(^\text{14}\) One-size-fits-all solutions and ‘best practices’ may simply be illusory if contextual factors are determinative of outcomes.

A separate theoretical issue is whether the concept is deemed to be an absolute good. The rule of law is so conceived: no one suggests that less rule of law is better than more. No one complains that that officials follow the rules too predictably, or that the legislature is too clear in announcing rules prospectively. On the other hand, individual components of the rule of law might not be absolute goods, but rather, goods for which we should think of in terms of optimal rather than absolute values. Judicial independence, for example, has been a major goal for reformers, who believe that only an independent and powerful judiciary can secure the rule of law. Yet, we have examples of societies that meet any definition of the rule of law, but do not feature a prominent role for the judiciary. Contemporary Japan provides a prominent example. The Japanese government structure has a well deserved reputation for following the rules and being free of corruption; the rule of law seems to be present. But the level of judicial autonomy is open to some

---


\(^\text{12}\) Thomas, ‘What Do The Worldwide Governance Indicators Measure?’, above n. 11, at p. 12.

\(^\text{13}\) Christiane Arndt and Charles Oman, ‘Uses and Abuses of Governance Indicators’, OECD Development Center Study (2006), at p. 51.

\(^\text{14}\) Thomas Carothers (ed.), Promoting the Rule of Law Abroad: In Search of Knowledge 2006, p. 139.
debate, with one prominent scholar characterizing it as being highly deferential to the ruling political party.\textsuperscript{15} We know further that too much judicial independence can lead to ‘juristocracy’ and the displacement of democratic authority. No one really thinks that the rule of law is identical to the rule of judges.\textsuperscript{16} Judicial independence is an example of a mid-level concept that is relatively clear, and about whose institutional manifestations there is some amount of consensus. But as a theoretical matter, we should not think that it is an unlimited good thing. Judicial accountability is also an important value, and indeed, we ought to think that a balance between independence and accountability is necessary for the rule of law to exist.\textsuperscript{17}

Note that if an overarching concept like the rule of law includes a sub-concept for which we can identify optimal, non-maximal values, we need to think carefully about relating the sub-concept to the larger one. We might think of the distance from the optimal value of judicial independence as corresponding to less rule of law; but problems of aggregating various concepts then come to the fore. If a particular jurisdiction has too much judicial independence and too little of another component (say prosecutorial capacity), aggregation might make it appear that \textit{on average}, it has the right level of each component, even though it fact it has neither. Aggregation might make the jurisdiction observationally equivalent to a country with optimal levels of both components.

We also need to be clear, as a theoretical matter, as to why we are choosing the concepts that we are choosing. What are the precise theoretical considerations that drive these choices? The centrality of corruption in current developmental discourse is an illustration of, potentially, choosing the wrong concept to focus on. Corruption has become a target of major institutional efforts to overhaul governance structures and change institutions. But in fact we have good reason to doubt that all the phenomenon that go under the label of corrupt behavior – defined as ‘misuse of public power for private gain’ – are really uniform.\textsuperscript{18} Economists have distinguished grand corruption from petty corruption, and some have argued that certain kinds of corruption can be ‘efficient grease’ given other institutional struc-

\textsuperscript{17}David Kosar, \textit{Judicial Accountability}, JSD Dissertation on file with author.
Corruption can increase the effective government wage, drawing marginally better people into the public sector. It is also the case that corruption seems endemic to certain stages of growth in many countries, and need not inhibit growth. Contemporary China provides one obvious example. But the United States also experienced deep corruption during the period between the Civil War and the rise of progressivism.

Corruption is treated in the current rhetoric as the proper object of regulation, yet its pervasive quality in many developing societies suggests that it may be playing some roles on which we do not have a good handle. As a theoretical matter, we are concerned with corruption not for its own sake, but because of its propensity to distort public decision making away from its efficient or socially desirable goals. If corruption is ubiquitous, we might see governments seeking to perpetuate inefficient policies because they facilitate enrichment of government officials. For instance, a country may impose restrictions on foreign trade as an inducement to both foreign and domestic firms to pay money under the table to evade those restrictions. A country may impose special restrictions on foreign companies doing business in the country, both as a way of encouraging domestic firms to pay large sums of money to keep out the foreign companies, as well as to encourage foreign firms to pay lots of money to buy their way in.

The point is that we are concerned with corruption because of its effects on the provision of public goods, and the implications for agency problems. Perhaps, then, the concept we should be trying to focus on is public goods provision. Corruption is, in some sense, an intervening concept that we think is correlated with what we are really concerned with, though perhaps, not identical. If we focus more directly on the real variable of interest, public goods provision, and treat corruption as proxy for it, we might bring into focus a new set of theoretical relationships between the two.

**How to Measure?**

Measurement is the subject of complex technical literature in various disciplines, and is a central concept in the philosophy of science. We evaluate the quality of measures in terms of validity (are they measuring what they purport to be measure-
Pitfalls of Measuring the Rule of Law

Efforts to measure the rule of law confront a number of particular challenges. Truly objective measures of institutional quality are difficult to construct, and surely this is the case for the rule of law. On the other hand, subjective measures are prone to concerns of validity and bias. To take one prominent example, many of the commercial risk analyses that measure governance rely heavily, or exclusively, on surveys of foreign investors. While these measures do a good job of capturing foreign investors’ views about the legal system or business environment, they do not necessarily capture the underlying reality of the quality of the legal system, and hence, may not be completely valid. Foreign investors may be biased in several ways: they may rely on limited information about the quality of the legal system; they may conflate other qualities of the country with the quality of the legal system; and most problematically, they may conflate a highly effective regulatory apparatus, which imposes effective constraints on environmental and labor practices, as poorly suited to doing business. In the latter case, the assessment would in fact be completely wrong: a high-quality legal system delivering policies disliked by investors would be rated poorly, though in fact it would be effective. It may also be that investors familiar with very corrupt environments are precisely the ones who have learned to navigate the maze of bribery demands, while more honest (and reliable) firms have been shut out of the market. In short, there are good reasons to think that business surveys do not produce particularly valid measures.

To be sure, there have been efforts to address these problems. The World Justice Project, surely the most ambitious project to attack the problem, recognized this problem and responded by including both elite and mass surveys about quality. A strong correlation among these components and with business surveys produced by other actors would increase confidence that the business surveys were not biased.

One common strategy for dealing with instances where we cannot measure law directly is to rely on proxy variables. One of the more creative efforts in this regard was an early one: the idea of using ‘Contract-Intensive Money’ as a measure of institutional quality. Contract-intensive money is roughly the percentage of the

---

26 M. Agrast et al., WJP Rule of Law Index, The World Justice Project.
money supply kept in formal institutions, such as banks. This is an excellent proxy for the quality of institutions, in that people are inclined to trust banks only when government regulation is of high quality, taxation systems are relatively clean, and the legal institutions can support and indeed create incentives for formal bank deposits.

Measurement error, however, has affected many projects purporting to measure law. This is particularly true of the concept of ‘Legal Origin’ associated with the team of La Porta, Lopez de Silvanes, Shleifer, and Vishny (LLSV), who have produced a host of studies purporting to show long-run impacts of the quality of legal institutions. Their studies look at the quality of corporate law, the structure of equity and debt markets, judicial quality, corruption, and many other variables, including economic growth. They find a consistent pattern that countries with French legal origin perform worse than those of English legal origin. This ties into old notions that the common law was a superior mode of regulation, because of its case-by-case incremental approach.28

The LLSV approach was hugely influential with the World Bank, which based policy advice on the findings. Unfortunately, the papers were riddled with measurement errors. First, take the very concept of legal origin. One of the central problems with the LLSV literature is its failure to disentangle the effects of legal origins from other institutional legacies associated with colonialism.29 In an important recent paper, a group of scholars attacked this problem by focusing on two set of countries: those which have ‘French’ law but were not colonized by France, and those which were British colonies but have so-called ‘mixed’ legal systems.30 It turns out that the growth rates in the former differ systematically from those in French colonies; while those in the latter are similar to those of other British colonies. Proxies for non-legal colonial policies (education, health) explain most of the observed difference in growth rates. The authors go on to examine the other dependent variables addressed by LLSV, including equity and debt markets, judicial quality, corruption, etc. in various multivariate specifications. They find that French colonies do better than non-French colonies with ‘French’

---


law, while British colonies of whatever legal tradition perform equally well. These results contradict those of LLSV by suggesting that colonial policies, rather than legal origin per se, is the source of differential performance.

The LLSV work also had systematic mis-codings. Holger Spamann re-analyzed the LLSV data on corporate governance and found coding errors in 33 out of 46 countries.\(^{31}\) This forced a major retroactive re-coding effort by the World Bank. Using his updated coding, the original LLSV claim that shareholder protection was higher in common law than civil law countries fell apart. Similarly, his paper on civil procedure found no systematic difference in the complexity or cost of civil procedure across categories, and also showed that the quality of German contract enforcement was superior to that of common law countries.\(^{32}\)

An even more fundamental critique of LLSV has to do with the measurement of the independent variable of interest, legal origin.\(^{33}\) LLSV divide the civilian legal tradition into a small number of families: French, German, and Scandinavian. They do so on the assumption that the key rules for the development of law are those governing private transactions, so that the distinct French and German civil code traditions ought to make a difference for outcomes. On this basis, for example, they code all of Latin America as having adopted French law, on the assumption that the Spanish tradition and independent borrowing led these countries to focus on the Code Civil as the main legal instrument.

This approach, however, ignores that many of the key institutions structuring law, including judicial independence, property rights, and limits on regulatory power, are typically found in national constitutions. Furthermore, patterns of constitutional origin, as opposed to civil code origin, are very different. Latin America, for example, was heavily influenced by North American ideas about constitutional design, and borrowed numerous institutions that are surely relevant to subsequent legal developments.\(^{34}\) One might then suggest that they enjoy, at least on this dimension, Anglo-American legal ‘origin’ as opposed to French. If this approach were taken, one would surely see the developmental advantages of the Anglo-American system disappear. The coding of the key variables, then, produced measures that were arguably invalid.


\(^{33}\) In addition to measurement error, the LLSV work has been subjected to numerous other critiques. See John Armour et al., ‘Law and Financial Development: What We Are Learning from Time-Series Evidence’, in: BYU Law Review 2009 (2009), pp. 1435-1500.

In short, measurement choices will influence outcomes. The fundamental concepts of validity, reliability, and bias are essential to remember when constructing measures, and the LLSV story surely argues for caution in translating results into policy recommendations.

**Can Measurement Undermine Improvements? Feedback Effects from Measurement Efforts**

A final pitfall, as near as I can tell unidentified in the literature, is that the provision of measures might, under some circumstances, actually undermine efforts to improve the behavior in question. Consider the widely reported Transparency International (TI) ratings, which ranks 180 countries on a ten point scale for levels of perceived corruption, aggregating several surveys. TI has done more than any organization to bring the issue of corruption to the center of the policy agenda. TI’s annual ratings receive wide coverage, and are utilized by governments and civil society in both high- and low-scoring countries. Singapore’s Economic Development Board, for example, trumpets its high score on the TI index, as well as its number one ranking on the Doing Business scale. On the other hand, the media in countries that score poorly on the survey are also wont to publicize the poor ranking. This is indeed part of the ‘name and shame’ strategy of TI.

The nature of corruption, however, makes it possible for the ‘name and shame’ strategy to backfire. The problem of corruption is essentially one of collective action on the part of the citizenry. Every citizen would be better off if the administration was honest and no one paid a bribe. On the other hand, if everyone else is paying bribes, an individual citizen is worse off if she fails to do so: she will not have access to the government services she is entitled to. A citizen’s expectations about the likely behavior of other citizens is the key factor in determining whether the collective action problem can be overcome.

If a citizen perceives corruption in their country to be a mild problem, but is then confronted with a sudden change in the externally reported assessment of the extent of corruption, the citizen will update her perceptions as to the likelihood of the problem. Movements in the index are not unheard of: Afghanistan fell from 117th in the world in 2005 to 179th in 2009. A study in the Ukraine found that people were much more willing to engage in corruption when they believed cor-

---

Pitfalls of Measuring the Rule of Law

Corruption was more widespread. Another way in which publicity might undermine the reality is that a perception of corruption might prevent new foreign investments from entering the country. This might cede the local market to those already in corrupt networks, preventing competition, and making it all the more difficult to move away from the corrupt equilibrium. Thus, publicizing corruption can perversely encourage it.

One might imagine a similar dynamic with regards to the rule of law. A low rating on an international indicator might be seen by both citizens and foreigners as a signal of actual quality of the underlying institutions. Citizens might then avoid using the courts, and may instead rely on alternative mechanisms to govern their transactions, such as limiting business to reputational networks. The lack of trust in the rule of law might also distort investment decisions, as people avoid investing in fixed assets subject to government predation or private appropriation. Foreigners might avoid investing in a country, knowing that they cannot rely on formal law and finding it difficult to break into local reputational networks. In short, as far as the rule of law goes, a bad image can become a self-fulfilling prophecy.

This is not a case for keeping measures secret. Far from it. But it does reiterate the importance of getting the measures right. Suppose, for example, that there is a country in which citizens, for reasons of culture or history, are systematically willing to report that levels of corruption are higher than they really are. This could be because citizens’ perceptions are wrong, or that there is a ‘culture of complaint’ whereby citizens are happy to criticize government officials. Giving wide voice to an inaccurate report of actual levels of corruption might induce those citizens who do think the government is trustworthy to weaken their faith. In this way, we can imagine that measures can undermine efforts to improve things.

Conclusion

We have identified some common pitfalls that need to be avoided in attempts to measure the rule of law. Some of these pitfalls are common to all social science research, and so grounding one’s efforts in the conventional rules thereof is a good place to start. These rules constrain one’s ability to measure something as abstract as the rule of law; but this does not mean that we should give up hope. Rather, we should renew our efforts to develop new measures that are reliable, valid, and with as little bias as possible. The World Justice Project effort is promising in this regard, but only further scrutiny will tell us if the efforts have been completely effective.

Our final pitfall, however, is about the use of measures, rather than the quality of measures themselves. Emphasizing the possibility of self-fulfilling prophecies may on its surface overstate the influence of obscure social science measures; yet at the same time it highlights that good governance is at the end of the day a coordination problem among citizens, in which one’s beliefs about other’s likely beliefs determine one’s actions.38
