doi:10.1017/aju.2026.10098

SYMPOSIUM ON MEASURING THE RULE OF LAW: CONTESTATION, POLITICS, AND POLICY

REACTIVE MEASURES: LESSONS FROM THE FIELD

*Kim Lane Scheppele**

Economist Charles Goodhart was considering how to measure the money supply when he said something that has passed into lore through the words of anthropologist Marilyn Strathern: "When a measure becomes a target, it ceases to be a measure."[1] Goodhart's law identifies a certain persistent problem in designing indicators. When a particular measure is known to those who are being measured and it is within their power to alter its value, they often will try to do so, thus ruining the usefulness of the measure.

We often think about measuring the rule of law as if those whose conduct is being observed were not watching us measuring them. But measuring social phenomena requires considering that people and institutions can and do react to being measured. As we think about building rule of law indicators, then, we should consider whether what we measure encourages good behavior or rewards deception. How we convert a messy on-the-ground assessment of rule of law into numbers should encourage rule of law to flourish rather than encouraging cheating.

I practice legal ethnography,[2] seeing legal systems whole rather than in disconnected variables, ripped out of context, so measuring discrete elements of a system across a large number of countries is not how I work. Ethnography is a close cousin of in-depth interviews, such as the ones explained by Janka Deli, Katerina Linos, and Mark Pollack in this symposium, teasing out how actors in specific times and places understand what they are doing.[3] That said, quantitative measurement is essential if one wants a quick overview of how countries differ in their commitment to the rule of law, something everyone from policymakers to academics may legitimately want to know.

First-rate quantification requires precise conceptualization, in addition to attention to ensuring validity, reliability, and the lack of bias.[4] Some of those issues can be fixed with care, but other issues cannot. When one "variabilizes" the world by cutting it up into discrete items to measure, one assumes that measuring A in one context is roughly equivalent to measuring the closest thing to A in another. But a prosecutor in Hungary does not look much like a prosecutor in the United States, so counting them in the same category already disregards key differences. In addition, the way A and B are connected in one case may not run through the same mechanisms as the way that A and B are connected in another case. A prosecutor may be established as a fourth branch institution in one country but part of the judiciary (or the justice ministry) in another.

So how can the knowledge we have gained from ethnographic research help to improve the way we count things? Among the many things that ethnographic knowledge can add to this picture, we need to consider that

* *Laurence S. Rockefeller Professor of Sociology and International Affairs, Princeton University, New Jersey, United States.*

[1] The history of this conversion is traced in C. Mattson, R.L. Bushardt & A.R. Artino Jr., *When a Measure Becomes a Target, It Ceases to be a Good Measure*. 13 J. GRAD. MED. EDUC. 2 (2021).

[2] Kim Lane Scheppele, *Constitutional Ethnography: An Introduction*. 38 L. & SOC. REV. 389 (2004).

[3] Janka Deli, Katerina Linos & Mark Pollack, *Interviews as a Method for Rule of Law Measurement*, 120 AJIL UNBOUND 339 (2026).

[4] Tom Ginsburg, *Thoughts on Measuring the Rule of Law*, 120 AJIL UNBOUND 300 (2026).

© The Author(s), 2026. Published by Cambridge University Press on behalf of American Society of International Law. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

people (and countries) will try to game the measures. Rather than proceed in the abstract, I will use my experience in post-communist Europe to illustrate the problem.

As first Hungary (starting in 2010) and then Poland (starting in 2015) began systematically destroying the independence of their judiciaries, the European Commission wanted some objective way to measure how far they had fallen and how they were different from the other member states. The EU Justice Scoreboard[5] was an early Commission tool to measure compliance with the rule of law, focusing on justice-sector performance.[6]

Up close, what was happening in those two countries was no mystery. In Poland, the government took a sledgehammer to the courts, packing the Constitutional Tribunal with unlawful appointments,[7] changing the system for appointing new judges from one that walled the judiciary off from politics to one that rewarded party loyalty.[8] (More than 3,000 judges were appointed in a politically compromised process over eight years.)[9] Polish courts routinely violated European Union and European human rights law under political tutelage.[10] The pro-autocratic government installed disciplinary procedures to punish judges who did not toe the party line.[11]

In Hungary, the process was more legal but just as destructive. The government expanded the number of judges on the Constitutional Court and changed the procedure for appointing them so that the governing party soon had a loyal majority.[12] The government created a National Judicial Office, directed by a party affiliate, to vet all new judges and to control the careers and salaries of judges already in the system.[13] The government changed the qualifications for judges currently on the Supreme Court to weed out those who might cause trouble and they changed the qualifications again when they had someone in mind to occupy the court presidency who did not meet the new qualifications.[14] Across the system, politically loyal court presidents were given the power to create panels of judges to whom they assigned specific cases so that the government could count on disputes of political interest always being assigned to friendly judges.[15]

The European Commission is very careful about treating all member states with equal respect, so it is reluctant to bring legal enforcement actions without hard evidence that the state is an outlier. In its first foray into coping with the "rule of law problem," the Commission developed the EU Justice Scoreboard to compare all countries across the EU, imagining that this would illuminate precisely where the independence of the judiciary had been compromised. Every year since 2013, then, every EU member state has dutifully filled out the long questionnaire to score themselves on various measures; other agencies around the EU have dutifully carried out requested

[5] Ylenia Guerra, *Measuring Justice? The EU Justice Scoreboard in the Light of the Performance-Based Approach*, *in* EU RULE OF LAW PROCEDURES AT THE TEST BENCH 157 (Christina Fasone, Adriano Dirri & Ylenia Guerra eds., 2024).

[6] András Jakab & Lando Kirchmair, *How to Develop the EU Justice Scoreboard into a Rule of Law Index: Using an Existing Tool in the EU Rule of Law Crisis in a More Efficient Way*, 22 GER. L.J. 936 (2021).

[7] WOJCIECH SADURSKI, POLAND'S CONSTITUTIONAL BREAKDOWN 132–49 (2019).

[8] Laurent Pech, *Poland's "Illegal Judges,"* VERFASSUNGSBLOG (Apr. 2, 2026).

[9] *Id.*

[10] Kim Lane Scheppele, *The Treaties Without a Guardian: The European Commission and the Rule of Law*, 29 COLUM. J. EUR. L. 93, 124–49 (2023).

[11] *Id.* at 134–37.

[12] Kriszta Kovács & Kim Lane Scheppele, *The Fragility of an Independent Judiciary: Lessons from Hungary and Poland – and the European Union*, 51 J. COMMUNIST & POST-COMMUNIST STUD. 189, 191–92 (2018).

[13] *Id.* at 192–93.

[14] Petra Bárd & Viktor Kazai, *Enforcement of a Formal Conception of the Rule of Law as a Potential Way Forward to Address Backsliding: Hungary as a Case Study*. 14 HAGUE J. RULE L. 165, 185–90 (2022).

[15] Hungarian Helsinki Committee, *A Brief Assessment of the Case Allocation Scheme and System of the Kúria Based on the Experiences of the Period Since the Entry into Force of the Judicial Reform* (Sept. 23, 2023).

surveys and aggregate data collection; and the European Commission has dutifully published the results in its annual Justice Scoreboard.[16] Producing the Scoreboard occupies many people for much of the year, each year.

How did the European Commission measure the rule of law in its Justice Scoreboard? Largely through tracking features of the legal system that were easy to count. They included measures of the efficiency of the judicial system: How many cases are there each year (and how many in each of a predefined set of legal categories)? How long does it take from beginning to end of a case, on average? How much of a backlog of cases is left at the end of each year?

Then, they included measures of the quality of the judiciary: What is the budget for the judiciary? How does the average judicial salary compare to the average national salary? What is the budget for legal aid services? What is the percentage of judges who are women? Have courts digitized their records? Are there offices in charge of enforcing judgments so that the courts have backup muscle?

Then, they included measures of the integrity and independence of the legal system: Does the general public perceive the courts as independent? Do companies? Are judges subject to sanctions if they fail to recuse in cases where they have conflicts of interest? How are prosecutors appointed? And who has the authority to fire them? Are disciplinary measures against lawyers subject to judicial review?

As it turned out, neither Hungary nor Poland scored badly on these measures. That is because the way that the judiciaries were captured tracked the stories I have told above, which intersected hardly at all with what was measured.

For example, the length of proceedings is not usually related to whether the judiciary has been compromised. A rule of law advocate should frown when it takes forever to resolve cases, as the Commission envisioned. But a rule of law advocate should also frown when a case takes almost no time at all. A trial with no evidence, affirmed by a judiciary that rubber stamps the decisions of lower courts can be especially speedy, which should constitute an even greater danger signal that judicial independence has been compromised. So, then, which end of this continuum gives you the "better" score rule of law? For what it is worth, in its 2026 Justice Scoreboard report, Hungary came in second among EU member states in having the speediest resolution of cases across all courts.[17] Poland was precisely in the middle. As Diego Zambrano notes in his contribution to this symposium, it is tragically easy to game many courtroom statistics like, for example, win rates in different kinds of cases.[18]

The Commission used budgets as a key indicator of the quality of the judiciary. Since poorer countries always do worse, standardizing the measure against some benchmark makes sense. I might have opted for the fraction of the national budget that goes to courts, but the Commission instead measured the expenditure on courts as a percentage of GDP. The top performers were Bulgaria, Latvia, and Malta with the worst scores going to Denmark, Cyprus, and Ireland.[19] Hungary and Poland, the rule-of-law-challenged countries, were squarely in the middle. When the European Network of Councils of the Judiciary interviewed judges across the EU to get their impressions of judicial independence, another potential measure of quality, the order was very nearly reversed.[20]

Or take another measure that tracks something different than what the Commission sought to discover, which was progress toward gender equality on the bench. The history of women judges is different in Eastern and

[16] All of the EU Justice Scoreboard Reports can be found at https://commission.europa.eu/strategy-and-policy/policies/justice-and-fundamental-rights/upholding-rule-law/eu-justice-scoreboard_en.

[17] European Commission, The 2026 EU Justice Scoreboard, COM(2026) 273, at 11, Fig. 6 [hereinafter 2026 EU Justice Scoreboard].

[18] Diego Zambrano, *Measuring the Rule of Law: How Often Do Governments Lose Important Cases?*, 120 AJIL UNBOUND 327 (2026).

[19] 2026 EU Justice Scoreboard, *supra* note 16, at 86, Fig. 34.

[20] European Network of Councils for the Judiciary (ENCJ), ENCJ Survey on the Independence of Judges 14 (2022).

Western Europe. In communist "Eastern Europe," virtually all of the judges were women because it was a low-status field, precisely because it was politically controlled.[21] When the EU asks now what proportion of the bench are women, it presupposes that the bench was male-dominated until EU gender equality rules kicked in so that a higher percentage of women is "good" because "progress." But according to the 2026 Justice Scoreboard, four of the top five countries on that measure are East-Central European countries, countries not otherwise known for their commitments to gender equality.[22]

I could criticize many of the other measures that the EU Justice Scoreboard uses—for example, keeping track of whether a procedure exists for disciplining judges who fail to recuse when they should—rather than asking whether judges who should have recused but did not actually faced consequences. Same with asking whether lawyers who were disciplined have recourse to courts rather than asking whether the procedure for disciplining lawyers is free of political control in the first place. (And, how would you gather data on this? Would you trust a government filling out the questionnaire to answer that question accurately?)

Because the Commission wanted to address outliers for the purposes of assistance and enforcement, it needed measures that could be used in all member states. That meant that countries doing one-off terrible things simply did not show up in the measures. The Commission did not ask, "Have judges been unlawfully appointed to the Constitutional Court?" or "Have new disciplinary procedures, staffed by politically appointed judges, been created?" or "Are judges disciplined for sending reference cases to the European Court of Justice?" Only Poland and Hungary would have answered yes, which was precisely why the questions were not asked.

In short, if you knew how judicial systems are corrupted these days from on-the-ground research, you would know that most of the measures that the Scoreboard uses are irrelevant at best. They simply do not capture how politicians compromise their judiciaries these days. It is as if someone who wanted to know the causes of crime measured the average height of buildings in the vicinity.

The Scoreboard goes beyond irrelevant to tragic, however, when Goodhart's law kicks in. Take the questions about legal aid, which were designed to measure access to justice. True, Hungary is the worst country in the EU on this measure in the 2026 Scoreboard Report,[23] but it would have been catastrophic on this measure if the Hungarian government did not pad the numbers. The Legal Aid Service in Hungary (*Jogi Segítségnyújtó Szolgálat*) does not just provide free legal counsel for court cases, but it also distributes identity documents, building permits, and more.[24] The entire budget of this agency was reported as "legal aid," even though it covers every public-facing legal service provided by government-employed lawyers.[25] Reporting "legal aid" this way disguises the pitiful provision of counsel for litigants, which is presumably what the Commission wanted to know. As the

[21] Barbara Havelková, Mathias Möschel & Śledzińska-Simon, *Women, Gender, and Constitutionalism in Central and Eastern Europe: Introduction to the Symposium*, 23 Int'l J. Con. L. 486, 491 (2025) ("It is true that women made up the majority of judges under state socialism and still do today in CEE [Central and East European] countries .... The state socialist judiciary had very limited power and weight, and CEE judiciaries have always been horizontally and especially vertically gender segregated, with the number of women in positions of power and at higher courts low, and declining in recent years.").

[22] 2026 EU Justice Scoreboard, *supra* note 16, at 87, Fig. 37. On commitments to gender equality across Europe, see Pew Research Center, Public Opinion Three Decades After the Fall of Communism, Ch. 7 Gender Equality (2019). While the huge gap on gender equality has narrowed between East-Central and Western Europe over those three decades, East-Central Europeans are still very much less feminist on issues like traditional gender roles in the household and whether men should take priority in scarce job markets.

[23] 2026 EU Justice Scoreboard, *supra* note 16, at 23, Fig. 22.

[24] International Legal Aid Group, *National Reports 2023: Hungary* 1 (click on Hungary).

[25] That said, the national law on legal aid permits legal aid services to be provided by civil society organizations or universities, if contracted through specialized lawyers registered as legal aid lawyers but that information has not been compiled in one place for reporting purposes. *Id.*, n. 7.

2023 country report based on interviews for the International Legal Aid Group noted about Hungary's legal aid system: "Unless there is willingness to invest in the system, it is likely to become completely dysfunctional."[26] The Justice Scoreboard does show that legal aid is a problem in Hungary but the magnitude of the problem is disguised. Failures of measurement like this may be one reason why Janka Deli concludes in her contribution that rule of law declines can happen much more quickly than statistics track because the statistics can mask serious problems for some time before a catastrophic collapse makes the underlying weaknesses visible.[27]

Take another measure that the Scoreboard deploys: digitalization of court records. Here, Hungary has one of the best scores in the EU. But the story behind the scenes is much more complicated. In Hungary, the government controls the courts by choosing court presidents who assign political cases to friendly judges.[28] In 2022, the European Commission withheld EU money that would otherwise have gone to Hungary but for its compromises of judicial independence[29] and demanded, among other things, that the Hungarian Supreme Court move to a system of randomized case assignment. The court obediently set up a system through which every case upon arrival was assigned a number in an electronic database that randomized the identifiers—before the case left the digitized system in order for the president of the Court to assign the case to friendly judges, at which point it was entered back into the digitized system.[30] The Court digitized the records (check!) but the corruption of the system remained.

I could go on, but you see the point. Things that are easy to measure are not necessarily what one needs to know to assess whether the rule of law is present or absent—and what can be measured that might in fact be useful can also often be gamed. Clever legalistic autocrats[31] have become masters of the one-off who-would-have-thought-of-it workarounds to avoid having indicators catch up with them.

Because legalistic autocrats invent new strategies for undermining the rule of law by constantly innovating, measures that rely on many countries being able to answer the same questions over time will miss most of what autocrats now do. So for example, when Viktor Orbán *lowered* the judicial retirement age in Hungary to remove many of the court presidents, the Justice Scoreboard would not have caught it because no one thought to measure stability of the retirement age as a factor in judicial independence.[32] Similarly when Viktor Orbán instituted a new qualification for a Supreme Court judge—that s/he had to have been a judge in the Hungarian legal system for five years before becoming eligible to sit on the Supreme Court—it looked like a sensible requirement, until one realized that it disqualified precisely one person: the Court president who had only sat on the Hungarian Supreme Court for three years, after he had been a judge at the European Court of Human Rights for seventeen years before that (which did not count).[33] Five years later, when the government wanted to appoint someone else as Supreme Court president, they amended the law to drop this requirement in order to parachute their favorite candidate into that position.[34] None of that made it into the Justice Scoreboard. And, in fairness

[26] *Id.*, n. 24.

[27] Janka Deli, *The Asymmetry Problem*, 120 AJIL UNBOUND 310 (2026).

[28] *See* Hungarian Helsinki Committee, *supra* note 15.

[29] Kim Lane Scheppele & John Morijn, *Frozen: How the EU Is Blocking Funds to Hungary and Poland Using a Multitude of Conditionalities*, VERFASSUNGSBLOG (Apr. 4, 2023). I might note that I worked on this problem for years as both scholar and advocate, writing academic articles, then reports for the European Parliament on why conditionality measures should be adopted. For the whole story, see Scheppele, *supra* note 10.

[30] Erika Farkas & András Kádár, *Trick and Treat?: Hungary's Game of Non-Compliance*, VERFASSUNGSBLOG (Dec. 12, 2023).

[31] Kim Lane Scheppele, *Autocratic Legalism*, 85 U. CHI. L. REV. 545 (2018).

[32] Kovács & Scheppele, *supra* note 12.

[33] Bárd & Kazai, *supra* note 15, at 185–89.

[34] *Id.* at 189–90.

to those who invented the indicators, who would have thought to measure things like these when they happened in precisely one country?

Knowing how legal systems work on the ground and in the field helps in understanding where measures of the rule of law can go wrong and how they may be improved. But between governments gaming the measures and engaging in behavior that evades capture, quantitative measures of the rule of law are always going to be a bit like the drunk who searches under the streetlight for the keys he dropped down the block. He may look where the light is, but it will not be where the keys are.