BRILL
NIJHOFF

EUROPEAN JOURNAL OF COMPARATIVE LAW AND GOVERNANCE (2023) 1–51

EJCL
brill.com/ejcl

Rule by Metrics

Performance, Quantification, and the Law

*Marta Infantino** | ORCID: 0000-0002-2876-8025
Associate Professor of Comparative Law, University of Trieste, Trieste, Italy
minfantino@units.it

Mauro Bussani | ORCID: 0000-0002-4213-9145
Full Professor of Comparative Law, University of Trieste, Trieste, Italy
Adjunct Professor, University of Macao, Macao, S.A.R. of the People's Republic of China
mauro.bussani@units.it

Received: 17 October 2023 | Accepted: 8 December 2023 |
Published online: 26 December 2023

Abstract

We live in societies in which an ever-growing array of activities in a range of fields are assessed, controlled and governed according to quantitative techniques. This paper focuses on processes grounded on quantitative tools (e.g. scores, indicators, rankings, algorithms) that are used in order to measure the performance of processes, people or organizations. Such processes are referred to in this paper as 'legal metrics'. We explore how different legal systems and sectors rely on legal metrics, and develop a

* This article draws on a research project on 'Legal Metrics: Quantification of Performances as a Regulatory Technique' carried out by the authors with Professor Sabino Cassese. Earlier drafts of the article were presented by the authors at the European Law and Governance School in Athens (Greece) on July 29, 2021 and on July 27, 2022, at the 21th General Congress of the International Academy of Comparative Law in Asunción (Paraguay), on October 25, 2022, and at the Scuola Superiore Meridionale in Naples (Italy), on December 15–16, 2022. The authors wish to thank Başak Başoğlu, Ivan Cardillo, Sabino Cassese, Giuditta Cordero Moss, Lukas Heckendorn Urscheler, Andrzej Jakubowski, Berk Kapancı, Augustín Parise, Teresa Rodríguez de las Heras Ballell, Nikolas Voulgaris, all the participants in the above mentioned meetings, and the anonymous reviewers of the EJCL for their insightful comments, as well as Thomas Roberts for the language editing. The usual disclaimers apply.

twofold argument. On the one hand, we argue that performance-based quantification often qualifies as a form of regulatory intervention. On the other hand, we contend that performance-based measures are always adopted and applied in legal contexts that react differently to the quantification turn. Both sides of this picture have largely been overlooked by comparative lawyers, even though they imply looking for law in extraordinary places and appraising a variety of legal manifestations in different settings. Understanding which forms of quantitative measures are widespread, in which sectors and regions, made by whom and producing what regulatory effects, is a research task that could (and in our opinion should) be of interest for all comparative law scholars.

Keywords

quantification of performances – legal metrics – comparative law – regulation – new public management – law & economics – nudging – global indicators – algorithmic law

1 Introduction

Over the last decades, a shift towards increased reliance on performance-based measures in order to improve management and to steer behavior towards desired goals has become noticeable in many areas and sectors, and increasingly more so since the so-called digital revolution.[1] This trend is global and pervasive: although to different extents and in different ways, we all live in quantified societies, in which an ever-growing array of activities in a range of fields—from education to health, from work to credit, from consumption to sustainability and justice—are in various ways assessed, controlled and governed according to quantitative techniques.[2]

One of the clearest illustrations of the pervasiveness of performance-based measurements that we all experienced passively (and, as teachers, actively) is the practice of grading students in schools. When students are large in number and diverse, grading offers a neutral and rapid tool for comparatively testing

1 Cf C. O'Neil, *Weapons of Math Destruction* (New York: Crown Books, 2016); A. Supiot, *La gouvernance par les nombres* (Paris: Fayard, 2015) (also published in English as *Governance by Numbers. The Making of a Legal Model of Allegiance*, transl. S. Brown (Oxford-Portland: Hart, 2017)); M. Power, *The Audit Society: Rituals of Verification* (Oxford: OUP, 1994).

2 See the authors quoted *supra* n 1.

masses of disparate learners. It is no coincidence that, as early as the seventh century CE, the Chinese empire based the imperial examination system for the recruitment of state officials on comparative assessments of candidates' works.[3] In the West, it was only from the nineteenth century onwards, when access to education in Europe started to reach gradually beyond political and religious élites, that the need arose for tools that could objectively and rapidly manage the increasingly large and varied crowd of students: it was at this point that grading was adopted.[4]

The example of student grades is telling for our purposes from many points of view. It shows that performance quantification can easily become normalized by those who are subject to and/or participate in it. Grading students is not the only possible approach to learning. Nevertheless, the quantification of student performance has rapidly become the most common technique for managing education, and is rarely objected to.[5] Students' grading vividly illustrates why social quantification is notoriously defined as a 'technology of distance': it is widely used in order to manage communities of strangers in which other tools for controlling behavior (in particular those based on intimate knowledge and personal trust) are better replaced by objectified and standardized methods of social governance.[6] Furthermore, it is clear from the example of grades that the choice in favor of quantification has broader consequences, many of which are unintended. Grading changes the set of incentives for students, shifting the focus of learning from achieving knowledge to getting good grades.[7] Moreover,

3 See M. Li, H. Liu, W.-C. Huang, 'A Historical Investigation and Classification Analysis of Seven Controversies on the Reform and Abolition of the 1300-year Chinese Imperial Examination System', *International Journal of Educational Research* 114 (2022) 1–9.

4 M. Strathern, 'From Improvement to Enhancement. An Anthropological Comment on the Audit Culture', *Cambridge Anthropology* 19 (1996/7) 1–21, at 4–5; N. Postman, *Technopoly. The Surrender of Culture to Technology* (New York: Vantage Books, 1993) 13, 139–140.

5 "To say that […] someone is a 7.2 on a sensitivity scale, or that this man's essay on the rise of capitalism is an A—and that man's is a C +—would have sounded like gibberish to Galileo or Shakespeare or Thomas Jefferson. If it makes sense to us, that is because our minds have been conditioned by the technology of numbers so that we see the world differently than they did" (Postman, *supra* n 4, 13).

6 This has been masterfully shown by A. Desrosières, *La politique des grands nombres. Histoire de la raison statistique* (Paris: La Découverte, 2000, 2nd edn) and T.M. Porter, *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life* (Oxford: OUP, 1995). See also, more recently, N. Couldry and U.A. Mejias, *The Costs of Connection. How Data Is Colonizing Human Life and Appropriating It for Capitalism* (Stanford: Stanford University Press, 2019) 122–151; A. Broome and J. Quirk, 'Governing the world at a distance: the practice of global benchmarking', *Review of International Studies* 41 (2015) 819–841.

7 Amongst many, see T.R. Guskey and S.M. Brookhart, 'Introduction', in T.R. Guskey and S.M. Brookhart (eds), *What We Know about Grading* (Alexandria: ASCD, 2019) 1–12, at 3.

once grades exist, they often take on a life of their own, meaning that a large plethora of actors outside the education system rely on them as a proxy for personal merit; after all, it is much easier to examine past grades than to test people's actual abilities.[8] Finally, the global practice of student grading also demonstrates that performance measurement manifests itself in forms and with features that differ significantly from one place to another, being heavily dependent on the context in which quantification occurs. As will be well-known to anyone who has had the opportunity to be educated, or to teach, in more than one country, the style, significance and practices of grading vary greatly across borders.[9]

This article is not about students' grades. What we rather aim to explore is how different legal systems, orders and fields rely on performance quantification as a regulatory tool. Processes grounded on quantitative techniques are referred to in this paper as 'legal metrics'. The argument we would like to develop is twofold.

On the one hand, performance quantification typically triggers intended and unintended behavioral changes in the processes and activities being measured, because human beings are well-known to react to measurement in co-reflexive ways.[10] Human reactivity to measurements is the main reason why performance-based quantification can qualify as a governance technique and as a form of regulatory intervention—that is, as one of the many elements (including politics, policies, formal and informal institutions, controls and compliance) that attempt to shape society.[11] As a tool that is evidence-based, mechanistic, and technical, quantification regulates the social world in

8 Guskey and Brookhart, *supra* n 7, 3; Langdon Winner, *Autonomous Technology. Technics-out-of-Control as a Theme in Political Thought* (Boston: MIT Press, 1977) 235.

9 C. Dilon, 'Grading Discrepancy in Global Education', *PEOPLE: International Journal of Social Sciences* 4 (2018) 1611–1624.

10 Human reactivity to measurements is undisputed in a number of disciplines—from sociology to psychology, from economics to anthropology. Cf H.A. Landsberger, *Hawthorne Revisited* (Ithaca: Cornell University Press, 1958) (discovering the 'Hawthorne effect', according to which individuals modify their behavior in response to their awareness of being observed); D.T. Campbell, *Assessing the Impact of Planned Social Change* (Hanover: The Public Affairs Center, 1976) (who, as a psychologist, noted that "[t]he more any quantitative social indicator (or even some qualitative indicator) is used for social decision–making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor": *Id.* at 49); C. Goodhart, 'Problems of Monetary Management: The U.K. Experience', in A.S. Courakis (ed), *Inflation, Depression, and Economic Policy in the West* (New York: Rowman & Littlefield, 1981) 111–132 (remarking, from an economic point of view, that "any observed statistical regularity will tend to collapse once pressure is placed upon it for control purposes": *Id.* at 116); Strathern, *supra* n 4 (noting, from an anthropological perspective, that "when a measure becomes a target, it ceases to be a good measure": *Ead.* at 5).

distinctive ways. It will be argued that performance measurement implies a shift in the techniques of governance, favoring overreliance on carrot-and-stick approaches as well as on form over substance.[12] When quantifying performance matters, the quality of performance becomes less important than the results that the actor measured can produce (and that the supervising authority can check).[13] Furthermore, the ubiquity of performance-based measures entails the emergence of new sets of methodologies and new competencies for intervening in the social world: counting requires people, offices and tools to collect data and monitor activities. Quantitative expertise thus becomes a factor of specialization as well as a source of technocratic legitimacy.[14]

On the other hand—and this is the second strand of our argument—we will show that, although globally widespread, rule by metrics is not a unitary phenomenon. Performance-based quantification comes in a variety of shapes and sizes, and permeates legal systems, orders and fields in different ways. It is, indeed, quite intuitive that grading students is not like (and does not raise the same implications of) publishing sovereign credit ratings, rewarding consumers' loyalty, reviewing a food delivery service, evaluating the quality of justice, or assessing human rights compliance within a supply chain.[15] It is perhaps less intuitive, but no less valid, that social performance quantification is permitted and accepted, and indeed occurs, in different legal settings—meaning both legal systems and inter-/trans-national orders—to different extents and in different forms.[16] In other words, performance-based

11 Needless to say, 'governance' is generally used as a broader term than 'regulation'. In the words of Braithwaite, Coglianese, and Levi-Faur, governance is "about providing, distributing, and regulating", while regulation "can be conceived as that large subset of governance that is about steering the flow of events and behavior, as opposed to providing and distributing": J. Braithwaite, C. Coglianese, D. Levi-Faur, 'Can regulation and governance make a difference?', *Regulation and Governance* 1 (2007) 1–7, at 3; see also A. Colombi Ciacchi and D. von der Pfordten, 'Exploring the relationship between law and governance: a proposal', *The Theory and Practice of Legislation* 11(2) (2023) 151–167, doi:10.1080/20508840.2023.2215657; R. Paul and M. Mölders, 'Introduction', in R. Paul, M. Mölders, A. Bora, M. Huber, P. Münte (eds), *Society, Regulation and Governance. New Modes of Shaping Social Change?* (Cheltenham: EE, 2017) 1–14, at 2–6; D. Levi-Faur, 'Regulation and Regulatory Governance', in D. Levi-Faur (ed), *Handbook on the Politics of Regulation* (Cheltenham: EE, 2011) 3–21, at 3; S. Cassese, 'Administrative Law Without the State? The Challenge of Global Regulation', *NYU Journal of International Law and Politics* 37 (2005) 663–694.

12 See section 6.2. The Dark Side.

13 See section 6.2. The Dark Side.

14 See section 6.2. The Dark Side.

15 For these examples, see *infra*, section 3. Instances of Legal Metrics.

16 See *infra*, section 3. Instances of Legal Metrics and 7. Comparative Law Matters.

quantification affects (and is affected by) legal systems, orders and fields in ways that are varied and context-dependent. Investigating how ruling by metrics works in a multiplicity of directions, depending on its style and context of application, is in our opinion just as important as acknowledging its regulatory potential. The spread of legal metrics may be global, but its (g)localized instantiations differ from one another and need to be appreciated from a comparative perspective.

With these objectives in mind, we will first define what we mean by 'legal metrics' (section 2), providing some examples of instances of the performance-based measures that we will consider in the paper (section 3). We will then survey the state of art of the legal literature both as regards the phenomenon as a whole, and also concerning some of its manifestations (section 4). Against the limited and fragmentary backdrop provided by current studies, section 5 will show where, how and for what purposes performance-based measurements are used. On this empirical basis, section 6 will explain what these performance-based measures have in common as a regulatory technique, while Section 7 will highlight the features in terms of which they differ and the many reasons why a comparative approach to the subject is much needed. Section 8 will summarize the results and suggest possible directions for future research.

2 The Meaning of Legal Metrics

The expression 'legal metrics' is rarely used in legal jargon. Nevertheless, the adoption of quantitative processes or yardsticks is required by law or entails legal consequences in a large number of fields. Consider, for instance: the rules on companies' accounting and public budgeting; finance, banking and insurance law; antitrust and tax law; rules on employees' incentives; technical standards for products and services; criteria for quantifying damages and calculating fines; practices of regulatory impact assessments; and evidence-based policy-making.

'Legal metrics' is used here with a different meaning. The term 'metrics' refers to the use of quantitative tools of any kind—points-based systems, black/white-lists, scores, outcome evaluations, indicators, rankings, algorithms—that continuously or repeatedly measure the performance (broadly conceived) of processes, people or organizations. Iteration is a fundamental feature of these tools: the threat/promise of measuring again in the future is what triggers the behavioral response by those who are measured (and in turn leads

to changes in the conduct of those who measure).[17] The adjective 'legal' refers to the direct or indirect effects that the act of measuring produces on the subjects involved in the measurement process, including not only those who are measured, but also those who make the measure and those who rely on the measurements. It goes without saying that we move in an uncharted territory, and that therefore the definition of 'legal metrics' herein provided is open to challenge and revision in light of new evidence and arguments. In the same vein, we intentionally keep the notions of performance, processes/people/organizations, and legal effects as broad as possible in order to avoid imposing any artificial limits on our search for social activities that are quantified, and hence regulated, through legal metrics.

When defined in these terms, legal metrics notably entails the application of quantitative techniques to areas that are traditionally not considered to be quantifiable, such as the quality of public administration, the strength of legal research, the efficacy of the law and the regularity of people's behavior. It is however important to stress that the 'legal' character of such metrics does not necessarily stem from the object being quantified, which may, or may not, be related to the law. Performance-based measurements sometimes have a clear legal focus, insofar as they assess the management of legal services or a person's compliance with legal norms.[18] However, they often involve non-legal fields, such as for instance the example of quantification initiatives in the broader area of education mentioned above.[19] We are interested in these initiatives too, and include them within our notion of legal metrics. This is because, by channeling expectations and conduct in given directions, they may also effectively exert regulatory influence.

17 On these effects, see *supra*, section 1. Introduction, as well as J.G. Kelley, *Scorecard Diplomacy. Grading State to Influence Their Reputation and Behavior* (Cambridge: CUP, 2017) 12–13, 18, 50, 246–247; M. Infantino, 'Global Indicators', in S. Cassese (ed), *Research Handbook on Global Administrative Law* (Cheltenham: EE, 2016) 347–367, at 348–349; A. Cooley, 'The emerging politics of international rankings and ratings. A framework for analysis', in A. Cooley and J. Snyder (eds), *Ranking the World. Grading States as a Tool of Global Governance* (Cambridge: CUP, 2015) 1–38, at 14; K.E. Davis, B. Kingsbury, S. Engle Merry, 'Introduction. The Local-Global Life of Indicators: Law, Power, and Resistance', in S. Engle Merry, K.E. Davis, B. Kingsbury (eds), *The Quiet Power of Indicators. Measuring Governance, Corruption, and Rule of Law* (Cambridge: CUP, 2015) 1–24, at 12, 19.

18 It is sufficient to consider the tools that focus on contractual performance by parties in the platform economy, or on respect by corporations for 'environmental, social and governance' (ESG) criteria and 'corporate social responsibility' (CSR) standards. For some concrete examples of quantitative initiatives of this kind, see *infra*, section 3. Instances of Legal Metrics.

19 Another widespread example concerns much of the automated tracking of online users' behavior. See *infra*, section 3. Instances of Legal Metrics.

For the same reason, the notion of legal metrics adopted in this paper covers performance-based measures that are markedly different from one another in terms of source and form. Quantification initiatives come in many different shapes and sizes, ranging from more qualitative, analog and expert-driven tools to hyper-quantitative, automated and AI-powered assessments. Legal metrics may be officially established by national and/or international law,[20] or may be developed and adopted by actors with limited to no official rule-making power.[21] Similarly, the legal effects of performance measures may be established by law, or may result from spontaneous reliance on them by the communities concerned.[22]

All the above implies that some of the quantitative tools analyzed in this paper satisfy all of the requirements of legality under a formalistic vision of the law, while the legal nature of others should be established by adopting a wider lens on the life of law, including its unofficial, soft and pluralistic manifestations.[23] We understand that embracing this view may run counter to deeply embedded ideas about the law.[24] Yet, looking for legal metrics requires

20 For instance, many measurements of judicial performance and of academic research are required under national laws: see *infra*, section 3. Instances of Legal Metrics. As regards international law, one of the most well-known examples is the obligation of states parties under Article 31 of the United Nations Convention on the Rights of Persons with Disabilities (CRPD) "to collect appropriate information, including statistical and research data, to enable them to formulate and implement policies to give effect to the present Convention".

21 Such as international organizations, non-governmental organizations (NGOs), domestic agencies, corporations. For some illustrations, see *infra*, section 3. Instances of Legal Metrics.

22 See *infra*, section 3. Instances of Legal Metrics.

23 On the necessary use of these lenses to investigate the contemporary dynamics of the law at any level—whether at global, transnational, domestic or micro-local level –, see P. Zumbansen, 'Transnational Law: Theories and Applications', in P. Zumbansen (ed), *The Oxford Handbook of Transnational Law* (Oxford: OUP, 2021) 3–30, especially at 15–22; B.Z. Tamanaha, 'A Reconstruction of Transnational Legal Pluralism and Law's Foundations', in N. Krisch (ed), *Entangled Legalities Beyond the State* (Cambridge: CUP, 2021) 449–477; P.S. Berman, 'Understanding Global Legal Pluralism. From Local to Global, from Descriptive to Normative', in P.S. Berman (ed), *The Oxford Handbook of Global Legal Pluralism* (Oxford: OUP, 2020) 1–35, at 2–12; M. Bussani, 'Strangers in the Law: Lawyers' Law and the Other Legal Dimensions', *Cardozo Law Review* 40 (2019) 3125–3184, at 3127–3160; G.T. Shaffer, 'Theorizing Transnational Legal Ordering', *Annual Review of Law and Social Science* 12 (2016) 231–253; S. Cassese, E. D'Alterio, M. De Bellis, 'The Enforcement of Transnational Private Regulation: A Fictitious Oxymoron', in F. Cafaggi (ed), *Enforcement of Transnational Regulation: Ensuring Compliance in a Global World* (Cheltenham: EE, 2013) 331–371, at 347–355, 367–368; L.M. Friedman, 'Erewhon. The Coming Global Legal Order', in L.M. Friedman, R. Pérez-Perdomo, M.A. Gómez (eds), *Law in Many Societies. A Reader* (Stanford: Stanford University Press, 2011) 308–316.

24 On these deeply embedded ideas about the law, see Bussani, *supra* n 23, 3128–3129.

searching for measures that perform regulatory functions independently of their status, and accepting that the law may be found in unusual places where lawyers usually do not venture.

3 Instances of Legal Metrics: A Selective Overview

There are countless possible examples of legal metrics as defined above.

In many countries, justice, education and health services are subject to domestic assessments intended to check their quality, to decide on the allocation of resources and to foster improvement.[25] At the same time, the performance of countries in the justice, education and health sectors is also

25 With regard to justice, see P.R. Borges Fortes, 'Quantification in State-administered Justice in Brazil', in M. Bussani, S. Cassese, M. Infantino (eds), *Comparative Legal Metrics. Quantification of Performance as a Regulatory Technique* (Leiden: Brill, 2023) 21–38 (Brazil); L.J. Béjar, J.A. Casanovas, C.A. Villanueva, 'Performance-based Evaluation in Mexico's Federal Administrative Tribunal and the Federal Judiciary Power: A Comparison', ibidem, 39–60 (Mexico); M.S. Sunder Raj and C. Basak, 'Judicial Performance Index in India: Charting a New Course', ibidem, 61–73 (India); A. Lienhard, 'Performance Assessment in Courts—The Swiss Case', ibidem, 74–95 (Switzerland); K.H. Ng and P.C.H. Chan, '"What Gets Measured Gets Done": Metric Fixation and China's Experiment in Quantified Judging', *Asian Journal of Law and Society* 8 (2021) 255–281 (China); J.C. Botero, A.M. Pinzon-Rondon, C.S. Pratt, 'How, When and Why Do Governance, Justice and Rule of Law Indicators Fail Public Policy Decision Making in Practice?', *Hague Journal of Rule of Law* 8 (2016) 51, 54–58 (Colombia); J.J. Spigelman, 'Measuring Court Performance', *Journal of Judicial Administration* 16 (2006) 69–80 (Australia). With regard to measurement by law schools, K.G. Weatherall and R. Giblin, 'Inoculating Law Schools Against Bad Metrics', in Kathy Bowrey (ed), *Feminist Perspectives on Law, Law Schools and Law Reform: Essays in Honour of Professor Jill McKeough* (Alexandria: The Federation Press, 2021) 191–223 (Australia); W. Nelson Espeland and M. Sauder, 'Rankings and Reactivity: How Public Measures Recreate Social Worlds', *American Journal of Sociology* 113 (2007) 1–40 (United States); on the measurement of legal scholarship (in both Europe and North America), R. van Gestel and A. Lienhard (eds), *Evaluating Academic Legal Research in Europe. The Advantage of Lagging Behind* (Cheltenham: EE, 2019); Rob van Gestel, 'Ranking, Peer Review, Bibliometrics and Alternative Ways to Improve the Quality of Doctrinal Legal Scholarship', in R. van Gestel, H.-W. Micklitz, E.L. Rubin (eds), *Rethinking Legal Scholarship. A Translatlantic Dialogue* (Cambridge: CUP, 2017) 351–398. With regard to education and health, J.Z. Muller, *The Tyranny of Metrics* (Princeton: Princeton University Press, 2018) 67–125 (United States); O'Neil, *supra* n 1, 50–67, 161–179 (United States). With regard to health, with a special focus on domestic quantification efforts during the Covid-19 pandemic, cf J. Liu, 'Social data governance: Towards a definition and model', *Big Data & Society* 1–14 (2022); C. Angiolini, 'Case Law Survey on Data Protection—Covid-19 Litigation Project', *Legal Policy and Pandemics* 1 (2021) 197–224; L. Bradford, M. Aboy, K. Liddell, 'COVID-19 contact tracing apps: a stress test for privacy, the GDPR, and data protection regimes', *Journal of Law and the Biosciences* 7 (2020) 1–21, doi:10.1093/jlb/lsaa034.

evaluated and compared by a plethora of transnational actors. A renowned instance of quantitative measurement in the field of justice is the body of judicial statistics published biennially by the European Commission for the Efficiency of Justice within the Council of Europe.[26] In the field of education, it is sufficient to recall the variety of tools produced by specialized institutions, companies and magazines for quantifying and ranking universities, scientific journals and individual researchers according to quality.[27] As to healthcare, alongside the myriad instances of performance assessment for doctors and the (self-)tracking of people's health through connected devices,[28] we may also consider the avalanche of performance-based metrics on countries' and regions' reactions to the spread of Covid-19, which have proliferated since the outbreak of the pandemic in 2020.[29]

Performance quantification is by no means limited to justice, education, and healthcare. The creditworthiness of sovereign states (and of companies and people) has long been measured by credit rating agencies, with the biggest share of the market being taken by the three American sisters Standard and

26 See coe.int/en/web/cepej/eval-tools (retrieved 7 December 2023).

27 The most famous include the 'Shanghai's Academic Ranking of World Universities' (shanghairanking.com (retrieved 7 December 2023)), launched in 2003 by the Center for World-Class Universities (CWCU) and the Graduate School of Education of the Shanghai Jiao Tong University, and now published by the Chinese independent agency Shanghai Ranking Consultancy; the 'Times Higher Education in the World University Rankings' (timeshighereducation.com/content/world-university-rankings (retrieved 7 December 2023)), published since 2004 by the British weekly magazine Times Higher Education; and the 'QS World University Rankings' (topuniversities.com (retrieved 7 December 2023)), published since 2009 by the US company Quacquarelli Symonds. The power of university rankings is confirmed by the recent scandal involving Columbia University. In the spring of 2022, a Columbia math professor, Michael Taddeus, denounced a pattern of inaccurate or misleading statements made by Columbia University to U.S. News and World Report in support of the national universities published in recent years (see math.columbia.edu/~thaddeus/ranking/investigation.html (retrieved 7 December 2023)). The denunciation prompted U.S. News and World Report to downgrade Columbia from number 2 to 18. See J. Barron, 'Columbia Loses A-Plus Status in U.S. News Rankings', *New York Times*, 14 September 2022, nytimes.com/2022/09/13/nyregion/columbia-loses-a-plus-status-in-us-news-rankings.html (retrieved 7 December 2023).

28 Both forms of measurement are particularly widespread in common law jurisdictions: cf Muller, *supra* n 25, 103–125 (on hospitals); S. Zuboff, *The Age of Surveillance Capitalism* (New York: Public Affairs, 2019) 200–206, 213; S. Mau, *The Metric Society: On the Quantification of the Social* (Cambridge: Wiley, 2019) 69–74, 151–153; D. Lupton, *The Quantified Self: A Sociology of Self-Tracking* (Cambridge: Wiley, 2016); O'Neil, *supra* n 1, 161–178 (all on the use of self-measurements, especially by tech and insurance companies).

29 On these global initiatives, see M. Infantino, 'Hazards and fallacies of social measurements: global indicators in the pandemic', *International Journal of Law in Context* 17 (2021) 168–185, doi:10.1017/S1744552321000264.

Poor's, Fitch and Moody.[30] Between 2003 and 2020, investment climates of all countries around the globe were (in)famously ranked by the World Bank's Doing Business Reports,[31] which is now about to be replaced by its successor, the Business Ready (B-Ready) project.[32] This takes us into a further area in which performance-based measures fluctuating between the transnational and the international sphere are disseminated: quantification initiatives focusing on the respective performance of countries vis-à-vis a wide array of issues—from protection of human rights to the rule of law, from the fight against corruption and human trafficking to sustainable development. Besides the WB's DB (and forthcoming B-Ready), notable illustrations are the 'Freedom in the World' and the 'Global Rights' index, launched respectively by the New-York based non-governmental organization (NGO) Freedom House and by the Brussels-based International Trade Union Confederation in order to score countries in terms of their respect for citizens' and workers' rights.[33] The 'Rule of Law Index®', published by the Washington-based NGO World Justice Project, evaluates how effectively states adhere to the rule of law,[34] while the 'Corruption Perceptions Index' (by the Berlin-based NGO Transparency International) and the 'Trafficking in Persons' reports (by the United States Department of State) quantify efforts made by states to combat corruption and human trafficking.[35] Like their predecessor Millennium Development

30 A. Naciri, *Sovereign Credit Rating: Questionable Methodologies* (London-New York: Routledge, 2017); R. Abdelal and M. Blyth, 'Just who put you in charge? We did: CRAs and the politics of ratings', in Cooley and Snyder (eds), *supra* n 17, 39–59; M. Bussani, 'Credit Rating Agencies' Accountability: Short Notes on a Global Issue', *Global Jurist* 10 (2010) 1–13.

31 See archive.doingbusiness.org/en/doingbusiness (retrieved 7 December 2023). Literature examining (and criticizing) the Doing Business Reports, their methodology and their effects is abundant. See M. Infantino, 'Quantitative Legal Comparisons: Narratives, Self-Representations and Sunset Boulevards', *Journal of International and Comparative Law* 6 (2019) 287–306; G. McCormack, 'Why 'Doing Business' with the World Bank May Be Bad for You', *European Business Organization Law Review* 19 (2018) 649–676; M.M. Siems, 'Taxonomies and Leximetrics', in J.N. Gordon and W.-G. Ringe (eds), *The Oxford Handbook of Corporate Law and Governance* (Oxford: OUP, 2018) 228–248; N. Garoupa, C. Gómez Ligüerre, L. Mélon, *Legal Origins and the Efficiency Dilemma* (London-New York: Routledge, 2016); G. Xu, *Does Law Matter for Economic Growth? A Re-examination of the 'Legal Origin' Hypothesis* (Brussels-Cambridge: Larcier-Intersentia, 2014); S. Deakin and K. Pistor (eds), *Legal Origin Theory* (Cheltenham: EE, 2012).

32 See worldbank.org/en/businessready (retrieved 7 December 2023).

33 See freedomhouse.org/report/freedom-world and globalrightsindex.org/ (retrieved 7 December 2023).

34 See worldjusticeproject.org/rule-of-law-index/ (retrieved 7 December 2023).

35 See transparency.org/en/cpi/ and state.gov/trafficking-in-persons-report/ (retrieved 7 December 2023).

Goals (MDGs), the Sustainable Development Goals (SDGs) developed by the United Nations steer and assess different countries' respective performances in achieving the quantitative targets set out in the General Assembly's 2030 Agenda.[36] It is also important to note the proliferation of self-reporting, benchmarking and third-party certification initiatives that loosely monitor compliance with human rights obligations and 'environmental, social and governance' (ESG) and 'corporate social responsibility' (CSR) standards by multinational corporations.[37]

However, performance quantification as a regulatory tool applies well beyond the level of the state administration and corporate conduct. Besides the case of private credit ratings mentioned above, within (and among) a number of jurisdictions both private and public actors employ both low-tech and high-tech techniques in various contexts—from service delivery to the management of punishments—in order to score the actual and prospective behavior of people, whether as consumers, criminals, or citizens in general.[38] Finally, performance metrics is a new vogue on digital markets.

---

36 S. McInerney-Lankford and H.-O. Sano, 'Human rights indicators in development: definitions, relevance and current trends', in S.P. Marks and B. Rajagopal (eds), *Critical Issues in Human Rights and Development* (Cheltenham: EE, 2021) 368–394; S. Fukuda-Parr, 'Sustainable Development Goals', in T.G. Weiss and S. Daws (eds), *The Oxford Handbook of the United Nations* (Oxford: OUP, 2018, 2nd edn) 764–779; S. Murthy, 'Translating Legal Norms into Quantitative Indicators: Lessons from the Global Water, Sanitation, and Hygiene Sector', *William and Mary Environmental Law and Policy Review* 42 (2018) 385–446; R. Buchanan, K. Byers, K. Mansveld, "What Gets Measured Gets Done': Exploring the Social Construction of Globalized Knowledge for Development', in M. Hirsche and A. Lang (eds), *Research Handbook on the Sociology of International Law* (Cheltenham: EE, 2018) 101–121.

37 This is also the way in which the United Nations Office of the High Commissioner for Human Rights (UNOHCHR) and the United Nations' Global Compact work: see ohchr.org/en/issues/indicators/pages/hrindicatorsindex.aspx (retrieved 7 December 2023) and unglobalcompact.org (retrieved 7 December 2023). On the performance quantification in these fields, which typically relies on self-quantification, cf L. Catá Backer and M.B. McQuilla, 'The algorithmic law of business and human rights: constructing private transnational law of ratings, social credit and accountability measures', *International Journal of Law in Context* 1–19 (2022); T. Morochovič and L.L. Reimers, 'Hidden in the Shades. Patterns of Entanglement within the Web of Corporate Social Responsibility Law', in Krisch (ed), *supra* n 23, 318–350; G. Auld and L.H. Gulbrandsen, 'Private Regulation in Global Environmental Governance', in R. Falkner (ed), *The Handbook of Global Climate and Environment Policy* (Cambridge: Wiley, 2013) 394–411; Cassese, D'Alterio, De Bellis, *supra* n 23, 337–339; AJ Rosga and M.L. Satterthwaite, 'The Trust in Indicators: Measuring Human Rights', *Berkeley Journal of International Law* 27 (2009) 253–315.

38 Cf N. Vardi, *Creditworthiness and 'Responsible Credit'. A Comparative Study of EU and US Law* (Leiden: Brill, 2022) and A. Adimi Gikay, 'The American Way—Until Machine

Tech companies, digital platforms, social media and sharing economy actors widely employ algorithmic scorings and measures for products, services, news, opinions, search results, and customers themselves for a variety of purposes.[39]

The list of performance quantification illustrations could be longer, but we believe the examples set out above are sufficient to establish how pervasive the phenomenon investigated in this paper is. As highlighted above, in section 1, the scope of quantification of the social world now covers issues—justice and healthcare, compliance with international law obligations, the appropriateness of private conduct—that were until recently considered in non-quantitative terms only. In these fields, the exercise of performance measurement, whether embedded in official law or not, has provoked both intended and unintended

---

Learning Algorithm Beats the Law?', *Case Western Research Journal Law of Technology and Internet* 12, issue 1 (2023), scholarlycommons.law.case.edu/jolti/vol12/iss1/3 (both on consumer credit scoring in the United States and Europe); C. Slobogin, *Just Algorithms. Using Science to Reduce Incarceration and Inform a Jurisprudence of Risk* (Cambridge: CUP, 2021) (on the use of algorithms to identify and punish individuals in the United States); M. Infantino and W. Wang, 'Challenging Western Legal Orientalism. A Comparative Analysis of Chinese Municipal Social Credit Systems', *European Journal of Comparative Law and Governance* 8 (2021) 46–85 (on Chinese social credit); D. Mac Síthigh and M.M. Siems, 'The Chinese Social Credit System: A Model for Other Countries?', *Modern Law Review* 82 (2019) 1034–1071 (on Chinese social credit); L. Catá Backer, 'Next Generation Law: Data-Driven Governance and Accountability-Based Regulatory Systems in the West, and Social Credit Regimes in China', *Southern California Interdisciplinary Law Journal* 28 (2018) 123–172 (on consumer scoring in the West and in China); Virginia Eubanks, *Automating Inequality: How High-tech Tools Profile, Police, and Punish the Poor* (New York: St. Martin Press, 2018) (on the use of rating programs for the management of social services in the United States).

39 Cf Zuboff, *supra* n 28; Mau, *supra* n 28; S. Umoja Noble, *Algorithms of Oppression. How Search Engines Reinforce Racism* (New York: NYU Press, 2018); S. Ranchordás, 'Online Reputation and the Regulation of Information Asymmetries in the Platform Economy', *Critical Analysis of Law* 5 (2018) 127–147; A. Btihaj (ed), *Metric Culture: Ontologies of Self-Tracking Practices* (Bingley: Emerald Insights, 2018); Lupton, *supra* n 28; F. Pasquale, *The Black Box Society* (Cambridge: Harvard University Press, 2015) 58–100; H. Masum and M. Tovey (eds), *The Reputation Society: How Online Opinions Are Reshaping the Offline World* (Boston: MIT Press, 2011). A similar trend, often dubbed RegTech and SupTech, is noticeable among domestic and supranational regulatory and supervisory authorities, especially in the banking and finance sector, which increasingly tend to rely on performance-based technologies to monitor norm compliance: L. Grassi and D. Lanfranchi, 'RegTech in public and private sectors: the nexus between data, technology and regulation', *Journal of Industrial and Business Economics* 49 (2022) 441–479; J.H. Hee Jung, 'RegTech and SupTech: The Future of Compliance', in Jelena Madir (ed), *FinTech: Law and Regulation* (Cheltenham: EE, 2021) 291–316; D. Restrepo Amariles and G. Lewkowicz, 'Unpacking Smart Law: How Mathematics and Algorithms are Reshaping the Legal Code in the Financial Sector', *Lex Electronica* 25 (2020) 171–185.

changes in management styles in the sectors concerned, in the institutional frameworks governing them, and in the behavior of the people involved. In other words, reliance on performance-based measurements has changed the way in which these sectors are regulated, whether officially or not.

4 The State of the Art in the Literature

The notion of legal metrics used in this paper is not limited to a single, well-defined field of studies. Distinct strands of legal scholarship have dealt with one or more instances of legal metrics as defined above, or with its methodological and foundational aspects. The most important lines within the literature deserve to be mentioned, in order to better contextualize this study and to clarify its added value.

In this regard, two preliminary caveats are in order. First, the emergence of legal metrics in the last decades has occurred in parallel with an intellectual and academic turn to quantitative methods that has affected all social sciences—including the law[40]—; this turn is particularly clear in the field of comparative constitutional law, in which large cross-national quantitative studies have proliferated in recent years.[41] Although these phenomena are clearly interconnected, we do not intend to explore here how social sciences in general and legal scientists in particular have progressively turned to quantitative methodologies. Our aim is different: we wish to identify and summarize the main lines of research on quantification that have either paved the way for or otherwise investigated experiments of legal metrics as defined above.[42]

Secondly, there is a burgeoning social science literature on our topic. Even if the wealth of management and behavioral studies on the effects of performance-based measures are here disregarded,[43] a number of research

40 See, for all, I.A. Moosa, *Publish or Perish. Perceived Benefits versus Unintended Consequences* (Cheltenham: EE, 2018). Somewhat paradoxically, as social research has increasingly been based on quantitative methods, its outputs have also increasingly been quantified and subjected to performance assessments by peers, universities, funders and journals.

41 On this turn, see, among many others, N. Petersen and K. Chatziathanasiou, 'Empirical research in comparative constitutional law: The cool kid on the block or all smoke and mirrors?', *International Journal of Constitutional Law* 19 (2021) 1810–1834; Ran Hirschl, *Comparative Matters: The Renaissance of Comparative Constitutional Law* (Oxford: OUP, 2014) 151–191.

42 See *supra*, section 2. The Meaning of Legal Metrics.

43 It is sufficient to type into any library database or book repository the two keywords 'performance' and 'improvement'.

studies on the disruption brought about by reliance on quantitative tools in domestic and transnational governance have been written in recent years from non-legal perspectives. The quantification of the social world through performance measurement has attracted the attention of historians,[44] anthropologists,[45] sociologists,[46] international relations theorists,[47] political scientists,[48] and economists.[49] Although this article stands on the shoulders of this previous research (among many other studies), we will not be concerned here with such developments, as we are mainly interested with what occurs in the legal domain.

We will in particular focus on three bundles of legal scholarship: the first is more or less closely linked to the spread of managerial and economic approaches to the law (section 4.1); the second looks at law-making through rankings and indicators in the global sphere (section 4.2); and the third explores the evolution of the law in our increasingly digitalized and algorithmic contemporary societies (section 4.3). We will then briefly assess what these strands of literature have in common and how they have contributed to shedding light on the regulatory effects of performance quantification (section 4.4).

4.1 *Economic-Oriented Legal Research*

It is well-known that, from the 1980s onwards, the emergence of the so-called New Public Management (NPM) paradigms, inspired by neo-classical economic theories, fueled the widespread adoption by governments of approaches characterized by deregulation, pro-market culture, and result-based

44 Cf Desrosières, *supra* n 6; Porter, *supra* n 6; M. Poovey, *A History of the Modern Fact. Problems of Knowledge in the Sciences of Wealth and Society* (Chicago: University of Chicago Press, 1998).

45 See on all points S. Engle Merry, *The Seductions of Quantification. Measuring Human Rights, Gender Violence, and Sex Trafficking* (Chicago: University of Chicago Press, 2015); C. Shore and S. Wright, 'Audit Culture Revisited: Rankings, Ratings and the Reassembling of Society', *Current Anthropology* 56 (2015) 421–444.

46 Cf S. Brayne, *Predict and Surveil. Data, Discretion, and the Future of Policing* (Oxford: OUP, 2021); Mau, *supra* n 28; Lupton, *supra* n 28; Buchanan, Byers, Mansveld, *supra* n 36; Espeland and Sauder, *supra* n 25.

47 See especially J.G. Kelley and B.A. Simmons (eds), *The Power of Global Performance Indicators* (Cambridge: CUP, 2020); Kelley, *supra* n 17.

48 See, among many, Eubanks, *supra* n 38; A. Broome, A. Homolar, M. Kranke, 'Bad science. International organizations and the indirect power of global benchmarking', *European Journal of International Relations* 24 (2018) 514–539; Broome and Quirk, *supra* n 6.

49 Cf Morten Jerven, *Poor Numbers. How We Are Misled by African Development Statistics and What to Do About It* (Ithaca: Cornell University Press, 2013); Power, *supra* n 1.

performance.[50] NPM paradigms called for the adoption of forms of performance benchmarking protocols that could be applied to public sector employees, and for the imposition of limits on the regulatory power of the administration through the adoption of governance arrangements that outsourced rule-making to the regulatees. This is known as 'performance-based' regulation': regulatory authorities do not specify what is required but rather set a performance target for regulatees and monitor the achievement of that target over time.[51]

In the meantime, the concurrent global rise of law-and-economics (L&E) has contributed to legitimizing the idea that legal rules should be understood not so much in terms of value-driven ought-to-be, but rather as a matter of costs and incentives, and that they should be assessed in the light of their ability to achieve expected outcomes and to induce changes in behavior.[52] Two well-known off-shoots from this view are worth recalling: on the one hand, the launch in the late Nineties of the so-called 'legal origins' project, which sought to measure quantitatively the extent to which countries' legal origins affect their economic performance,[53] and, on the other hand, the development of behavioral L&E research on how non-regulatory measures, such as nudges, smart disclosures and choice architecture, influence people's behavior and could be used by regulators to channel social conduct in desired directions.[54]

---

50 M. Shamsul Haque, 'New Public Management: Origins, Dimensions, and Critical Implications', in K.K. Tummala (ed), *Public Administration and Public Policy* (Paris: EOLSS, 2004) 209–229, at 210–211.

51 See C. Coglianese, 'Performance-based regulation: concepts and challenges', in F. Bignami and D. Zaring (eds), *Comparative Law and Regulation. Understanding the Global Regulatory Process* (Cheltenham: EE, 2016, reprinted 2018) 403–429; C. Coglianese, J. Nash, T. Olmstead, 'Performance-Based Regulation: Prospects and Limitations in Health, Safety, and Environmental Protection', *Administrative Law Review* 55 (2003) 705–729. However, performance-based regulation has also been advocated for other purposes, e.g., as a tool for protecting consumers in post-industrial economies: cf L.E. Willis, 'Performance-Based Consumer Law', *University of Chicago Law Review* 82 (2015) 1309–1409; S.D. Sugarman, 'Salt, High Blood Pressure, and Performance-Based Regulation', *Regulation and Governance* 3 (2009) 84–102.

52 For a summary of these shifts, see L.A. Kornhauser, 'The Normativity of Law', *American Law and Economics Review* 1 (1999) 3–25.

53 See, e.g., R. La Porta, F.C. Lopez de Silanes, A. Shleifer, R.W. Vishny, 'Legal Determinants of External Finance', *Journal of Finance* 52 (1997) 1131–1150; R. La Porta, F.C. Lopez de Silanes, A. Shleifer, R.W. Vishny, 'Law and Finance', *Journal of Political Economy* 106 (1998) 1113–1155; E.L. Glaeser and A. Schleifer, 'Legal Origins', *Quarterly Journal of Economics* 107 (2022) 1193–1229 (2002).

54 This is the very well-known and influential thesis of R.H. Thaler and C.R. Sunstein, *Nudge. Improving Decisions About Health, Wealth and Happiness* (New Haven: Yale University Press, 1st edn 2008, 2021).

Leaving aside the actual ebbs and flows of such a *Weltanschauung*, the key point for our purposes is that these economic-inspired modeling approaches have made a powerful contribution to the creation of an intellectual milieu in which performance-based assessments are normalized as a core strategy for social and legal intervention.

4.2 *Global Indicators*

In contrast from economic-centered approaches, two other strands within the literature have considered the quantification of the social from an external perspective, investigating the regulatory implications of performance-based practices in respectively global arenas and the everyday digital world. Let us start with the former line of inquiry.

As noted above, in 2003 the World Bank launched the Doing Business Reports, which translated the legal origins theory into a global indicator of countries' business friendliness.[55] Following their launch, the evident limitations of the Reports led many legal scholars to criticize them and their regulatory effects.[56] Some even worked on alternative projects. For instance, frustration with the Doing Business Reports inspired three (at that time) Cambridge-based scholars to develop the 'CBR Leximetric Datasets', scoring countries for labor regulation and shareholder protection.[57] A similar sentiment led the French 'Fondation pour le droit continental' to inaugurate the 'Index de la sécurité juridique' (published twice, in 2015 and 2018), aiming to compare countries' degrees of 'legal certainty'.[58] The experience with the Doing Business team induced Juan

55 See n 31.

56 See the authors *supra* n 31. See also the articles collected in the special issues on 'Symposium on Legal Origins', *American Journal of Comparative Law* 57(4) (2009) 765–876, 'Economics and Comparative Law', *University of Toronto Law Journal* 59(2) (2009) 179–235, 'Law and Finance', *Brigham Young University Law Review* (2009) 1413–1906, 'Jurimetrics', *Journal of Institutional and Theoretical Economy* 166(1) (2010), 'Misurare il diritto', *Annuario di diritto comparato* (2012) 7–353, as well as M. Infantino, *Numera et impera. Gli indicatori giuridici globali e il diritto comparato* (Milano: Franco Angeli, 2019) 145–159; M.M. Siems, 'Legal Origins: Reconciling Law and Finance and Comparative Law', *McGill Law Journal* 52 (2007) 55–81; Association Henri Capitant des amis de la culture juridique francaise, *Les droits de tradition civiliste en question. A propos des rapports Doing Business* (Paris: Société de législation comparée, 2006, 2 vols); Bertrand du Marais (ed), *Des indicateurs pour mesurer le droit? Les limites méthodologiques des rapports Doing Business* (Paris: La Documentation française, 2006).

57 See repository.cam.ac.uk/handle/1810/256566 (retrieved 7 December 2023).

58 The ISJ, which was written in French and remained almost unknown within the global debate, was discontinued after the second edition in 2018: see B. Deffains and M. Séjan (eds), *Index de la sécurité juridique. Rapport pour la Fondation pour le droit continental* (Paris: Dalloz, 2018). For some critical remarks on this project, see N. Genicot, 'L'index

Carlos Botero to conceive the 'Rule of Law Index®', which has been published since 2008 within the 'World Justice Project' and is sponsored, *inter alia*, by the American Bar Association.[59]

The multiplication of global indicators has attracted the attention of a growing number of legal scientists, especially since the end of the 2000s. On the one hand, scholars interested in global governance, global administrative law and global legal pluralism have studied in particular the increased use of global quantitative tools for data collection in inter- and trans-national relations as instruments for establishing and enforcing normative standards.[60] On the other hand, scholars from other fields, such as law-and-development, comparative law, and human rights law, have emphasized the many unintended consequences that the growing reliance on quantitative, performance-based tools for controlling performance often precipitate in their respective domains.[61]

---

de la sécurité juridique, ou comment promouvoir le droit continental par le biais d'un indicateur', *Droit et société* 1 (2020) 211–234; J. Knetsch, 'Measuring Legal Certainty? Critical Feedback about the Development of an Index of Legal Certainty', in M. Fenwick, M.M. Siems, S. Wrbka (eds), *The Shifting Meaning of Legal Certainty in Comparative and Transnational Law* (Oxford-Portland: Hart, 2017) 177–186.

59 See worldjusticeproject.org; but see also Botero, Pinzon-Rondon, Pratt, *supra* n 25.

60 From the perspective of global governance, see D.V. Malito, G. Umbach, N. Bhuta (eds), *The Palgrave Handbook of Indicators in Global Governance* (Cham: Palgrave MacMillan, 2018). From the perspective of global administrative law, see Merry, Davis, Kingsbury (eds), *supra* n 17; K.E. Davis, A. Fisher, B. Kingsbury, S. Engle Merry (eds), *Governance by Indicators. Global Power through Quantification and Rankings* (Oxford: OUP, 2012) (and in particular the chapter of S. Cassese and L. Casini, 'Public Regulation of Global Indicators', 465–474). From the perspective of global legal pluralism, see D. Nelken, 'The Legitimacy of Global Social Indicators: Reconfiguring Authority, Accountability and Accuracy', *Cahiers de Droit* 59 (2018) 35–84; D. Restrepo Amariles, 'Legal indicators, global law and legal pluralism: an introduction', *Journal of Legal Pluralism and Unofficial Law* 47 (2015) 9–21; B. Frydman and A. Van Waeyenberge (eds), *Gouverner par les standards et les indicateurs. De Hume aux rankings* (Brussels: Bruylant, 2014).

61 From the law-and-development perspective, see S.P. de Souza, *Designing Indicators for a Plural Legal World* (Cambridge: CUP, 2022); M. Trebilcock and M. Mota Prado, *What Makes Poor Countries Poor? Institutional Determinants of Development* (Cheltenham: EE, 2011); from the perspective of comparative law, see Infantino, *supra* n 56; M. Versteeg and T. Ginsburg, 'Measuring the Rule of Law: A Comparison of Indicators', *Law and Social Inquiry* (2017) 100–137; from the human rights perspective, S. Walker, 'Challenges of human rights measurement', in B.A. Andreassen, H.-O. Sano, S. McInerney-Lankford (eds), *Research Methods in Human Rights. A Handbook* (Cheltenham: EE, 2017) 306–332; S. Fukuda-Parr, *The MDGs, Capabilities, and Human Rights: The Power of Numbers to Shape Agendas* (London-New York: Routledge 2015).

4.3 *Algorithms and the Law*

A more recent strand of legal scholarship has focused mainly on the changes to and challenges for legal systems brought about by mass digitalization and datafication.

The starting point for such research has been the consideration that, in (wealthy) contemporary societies, social activity and human behavior are increasingly tracked, quantified and managed by powerful private and public actors through technology.

This is why some scholars, especially from the United States and Europe, have investigated the potential of algorithmic metrics for profiling individuals and automating the personalization of legal rules. Discussions have considered how increasingly smart algorithms can be devised in order to provide people with ex ante, context-dependent behavioral prescriptions, or with individually-appropriate levels of information and disclosure, or with personally-tailored suggestions as to which standard of care and conduct should be adopted.[62]

Nevertheless, most studies on digitalization, algorithms and the law have adopted a rather more critical stance concerning the wider consequences of automating performance quantification. Some have examined the drawbacks associated, in selected jurisdictions, with the use of algorithmic quantification and automation by police, courts, governments and private companies.[63] Other scholars have investigated and denounced how, through choice architecture, big data analytics and practices of (self-)measurement, algorithmic coding can constrain and regulate social behavior, reducing decision-making autonomy

62 Cf C. Busch, 'Implementing Personalized Law: Personalized Disclosures in Consumer Law and Data Privacy Law', *University of Chicago Law Review* 86 (2019) 309–331; P. Hacker, 'Personalizing EU Private Law: From Disclosures to Nudges and Mandates', *European Review of Private Law* 25 (2017) 651–677; O. Ben-Shahar and A. Porat, 'Personalizing Negligence Law', *New York University Law Review* 91 (2016) 627–688. However, for a more critical perspective, see T. Endicott and K. Yeung, 'The death of law? Computationally personalized norms and the rule of law', *University of Toronto Law Journal* 72 (2021) 373–402; D.L. Burk, 'Algorithmic Legal Metrics', *Notre Dame Law Review* 96 (2021) 1147–1203; J.M. Barry, J.W. Hatfield, S.D. Kominers, 'To Thine Own Self Be True? Incentive Problems in Personalized Law', *William and Mary Law Review* 62 (2021) 723–790.

63 Cf T. Sourdin, *Judges, Technology and Artificial Intelligence. The Artificial Judge* (Cheltenham: EE, 2021); A. Guthrie Ferguson, *The Rise of Big Data Policing Surveillance, Race, and the Future of Law Enforcement* (New York: NYU Press, 2019); M. Veale and I. Brass, 'Administration by Algorithm? Public Management Meets Public Sector Machine Learning', in K. Yeung and M. Lodge (eds), *Algorithmic Regulation* (Oxford: OUP, 2019) 119–149; B. McGurk, *Data Profiling and Insurance Law* (Oxford: Bloomsbury, 2018).

and leaving people vulnerable to digital nudging.[64] Both kinds of studies emphasize the many weak spots that affect big data analytics and algorithmic automation, such as proneness to bias, inherent opaqueness and limited (if any) accountability.

4.4 *Common Threads*

Overall, the above lines of research should be credited for shedding light on how traditionally non-regulatory tools, such as nudges, indicators and algorithms, may be used to shape and channel individual and collective behavior.

While economic-oriented legal research has contributed to normalizing quantitative and outcome-driven approaches to legal rules, studies on the normative power of indicators and algorithms have further helped single out how performance-based, quantitative tools have an impact on (and are impacted by) the legal contexts in which they are adopted and deployed. Reliance on performance-based quantitative tools may for instance highlight correlations that would not otherwise be visible, and also simplify procedures, reshape processes and competencies for dealing with social issues, inform policy-making processes, inspire legal reforms, promote best practices, and stimulate worthy legal behavior.[65]

Nonetheless, as the above literature has also emphasized, performance-based quantitative tools are exposed to many hazards. They may suffer from methodological weaknesses (such as the entrenchment of their producers' biases), empower unaccountable technocratic elites with substantial and unleashed normative power, divert social attention from what counts towards what is counted and produce more and less dangerous unintended effects (including the inducement of rank-seeking and gaming strategies by those who are measured) that may frustrate or run counter the very purpose of the measurement.[66]

64 See, in addition to the critical voices mentioned *supra* n 63, P.R. Borges Fortes, P.M. Baquero, D. Restrepo Amariles, 'Artificial Intelligence Risks and Algorithmic Regulation', *European Journal of Risk Regulation* (2022) 357–372; H. Nowotny, *In AI We Trust: Power, Illusion and Control of Predictive Algorithms* (Cambridge: Wiley, 2021); K. Yeung, 'Hypernudge? Big Data as a Mode of Regulation by Design', *Information, Communication and Society* 20 (2016) 118–136; M. Hildebrandt, *Smart Technologies and the End(s) of Law: Novel Entanglements of Law and Technology* (Cheltenham: EE, 2015); D. Keats Citron, 'Technological Due Process', *Washington University Law Review* 85 (2008) 1249–1313; Lawrence Lessig, *Code and Other Laws of Cyberspace, Version 2.0* (New York: Basic Books, 2006).

65 See especially the authors cited *supra* n 56 and 60–62.

66 See the authors cited *supra* n 56 and 60–64.

However, one limitation affecting existing legal research on the transformative impact of the quantification of social phenomena is that it has adopted so far a limited scope. The above research has typically focused on reliance on specific quantitative techniques (e.g., indicators, ratings, algorithmic scoring) in specific sectors (e.g., human rights measures, finance and corporate law, business-to-consumer digital transactions), examining the implications of such reliance either in selected jurisdictions or in highly decontextualized ways. The same legal scholarship has also failed to notice the common thread that keeps together all manifestations of ruling by numbers.

5 The Features of Legal Metrics

Within this scientific context, this article aims to shed light on the overall legal consequences produced by increasing reliance on performance-based assessments, whether low- or high-tech, domestic or supranational, official or unofficial. We will focus on this in section 6. At the same time, another original contribution that this article aims to make to the debate on social quantification lies in its emphasis on difference: performance-based measures and their legal impact are not inherently uniform across systems, cultures and sectors. While further research could achieve more comprehensive descriptions of performance-based measures and their regulatory potential, we will argue in section 7 that legal metrics need to be subjected to comparative study on account of the variety of forms and impacts.

Before we get into this, it is possible, and in our view necessary, to clarify how legal metrics actually works. In fact, despite the abundance of forms that performance-based quantification may take, there are many common features that need to be unveiled. In this section, we will try to highlight where instances of legal metrics can be found (section 5.1), how, by whom and for whom they are made (sections 5.2–5.3). The last two sub-sections will investigate the reasons why legal metrics are deployed (section 5.4) and the consequences this entails (section 5.5).

5.1 *Where Metrics Can Be Found*

Performance-based measurements are on the rise everywhere, whereby 'everywhere' means any possible locus of law-making.

Many of the examples surveyed in section 3 concern instances of legal metrics adopted at the domestic level, either by state organs (through law or by ministries and agencies working in a regulatory capacity)[67] or by other actors, such as company-led social credit initiatives in China.[68] However, legal metrics are also often requested and produced by international organizations, both regional (such as the Council of Europe) and global (e.g., the World Bank)[69]. Still other forms of performance measurements flourish across, and independently from, national borders and international law regimes, in what may be called the transnational level of law-making.[70] This is for instance the case for assessments of corporations' ESG/CSR obligations[71] and algorithmic-based measurements of the online behavior of market players (whether businesses or consumers).[72]

More data would be needed in order to understand whether there are any patterns within the adoption of specific types of performance-based measurements in given regions and spheres. It is difficult to infer any generalizations not only because the available empirical data are limited, but also because any attempt at identifying certain fields as being more prone to quantification than others would have to be wary of the blurred lines and different understandings of the same domains across legal cultures. For instance, a correlation between the spread of quantification in some sectors and their public (or private) character would be largely meaningless. Areas that in some places are deemed to pertain to the core of the state's non-delegable duties may elsewhere be left in private hands, or in still other places may be managed in ways that defy classification according to the public/private dichotomy. Any attempt at generalization on the basis of the 'public-private' divide would fail on account of the different ways in which the divide is perceived and managed around the world.[73]

---

67 This is the case for national initiatives to assess the performance of courts, hospitals and education, as well as administrative bodies or the state itself. See section 3. Instances of Legal Metrics.

68 See section 3. Instances of Legal Metrics.

69 See the illustrations of the CEPEJ and of the DB mentioned in section 3. Instances of Legal Metrics.

70 Zumbansen, *supra* n 23, 3–30; Bussani, *supra* n 23, 3146–3160; Catá Backer, *supra* n 38, 123–172; Friedman, *supra* n 23, 308–316.

71 On this point, see, in addition to the authors *supra* n 37, E. Webster and L. Mai, 'Transnational environmental law in the Anthropocene', Transnational Legal Theory 11 (2020) 1–15; Auld and Gulbrandsen, *supra* n 37, 394–411.

72 See the authors quoted *supra* n 62–64.

73 Aside from the fact that the public/private is difficulty to reconcile with China's socialist market economy, one need only recall here that legal education and healthcare services

What we can and should underline is that measures existing in different places and on different levels rarely operate in isolation from one another. On the contrary, domestic, international and transnational initiatives constantly interact, compete and conflict with one another, and are also strengthened or influenced by one another. Let us try to consider some examples. Performance-based assessments of judicial performance have developed in parallel at the international and at the national level, sometimes in opposition and sometimes as complimentary to one another.[74] The transnational quantification of scientific output by companies providing metrics on journal and researcher performance (such as Scopus, Web of Science and Google Scholar[75]) is increasingly relied on by domestic, state-led assessments of universities, thereby pushing national academics to adopt strategies that reinforce the significance of transnational, corporate measurements.[76] In the field of global indicators, over their almost twenty years of existence World Bank (WB) Doing Business reports (DB) have gently induced governments around the world to adopt slightly under 4,000 national reforms of business law,[77] and have

---

are largely privatized in common law countries, while they are generally publicly funded and managed elsewhere. On the difficulties arising out from the application of the public/private divide across different cultures, see A. Bradford, *Digital Empires. The Global Battle to Regulate Technology* (New York: OUP, 2023) 69–104; Couldry and Mejias, *supra* n 6, 54–57; X. Yu, 'State Legalism and the Public/Private Divide in Chinese Legal Development', *Theoretical Inquiries in Law* 15 (2014) 27–52; Cassese, D'Alterio, De Bellis, *supra* n 23, 342–347; I. Castellucci, 'Rule of Law with Chinese Characteristics', *Annual Survey of International and Comparative Law* 13 (2007) 35–92, at 69–75; J.H. Merryman, 'The Public Law-Private Law Distinction in European and American Law', 17 *Journal of Public Law* 17 (1968) 3–19.

74 See the authors *supra* n 25.

75 See, respectively, scopus.com/sources (retrieved 7 December 2023), clarivate.com/webofsciencegroup/solutions/web-of-science (retrieved 7 December 2023) and scholar.google.com/intl/it/scholar/metrics.html (retrieved 7 December 2023). On the overall effects of the metrics provided by these companies in the field of research, see, among many, S. Lamdan, *Data Cartels* (Stanford: Stanford University Press, 2023) 52–55, 63–66; M. Biagioli and A. Lippman (eds), *Gaming the Metrics. Misconduct and Manipulation in Academic Research* (Boston: MIT Press, 2020); Moosa, *supra* n 40; A. Vinokur, 'La normalisation de l'université', in Frydman and Van Waeyenberge (eds), *supra* n 60, 235–261. Similar observations apply to the many private actors producing global rankings of academic institutions, such as THE, Quacquarelly Symonds and ARWU mentioned *supra* note 27; on these rankings, see B.M. Kehm, 'Global University Rankings: Impact and Applications', in Biagioli and Lippman (eds), *Gaming the Metrics*, this n, 93–100.

76 See the authors *supra* n 25.

77 World Bank, *Doing Business 2020. Comparing-Business-Regulation-in-190-Economies* (Washington D.C.: World Bank, 2019) 25, archive.doingbusiness.org/en/doingbusiness (retrieved 7 December 2023).

inspired many other experiments with legal indicators, both by the WB itself[78] and by others.[79]

The case of the DB shows well that tools for quantifying social phenomena are reactive to the environment in which they operate, and therefore often change through the repetition of their measurement cycles. For instance, the strong resistance voiced by France after early editions of the DB (suggesting that legal systems whose origins could be traced back to the British common law were more efficient than those whose origins lay in French law) prompted the WB to gloss over the significance of legal origins in later reports.[80] In 2009, harsh criticism by the International Trade Union Confederation and the International Labour Organization against the dimension 'Employing Workers' in the DB (which rewarded countries where firing workers was easier) obliged the WB to stop using it within the calculation for the final 'Ease of Doing Business' score.[81] Vociferous critiques from many quarters accusing the DB of

78 The DB reports inspired the creation of the 'Investing Across Borders' (IAB) index (published once in 2010 to measure countries' openness to foreign direct investments and then discontinued: see World Bank, *Investing Across Borders 2010* (Washington D.C.: World Bank, 2010), openknowledge.worldbank.org/handle/10986/27883 (retrieved 7 December 2023)), the 'Women, Business and the Law' (WBL) reports (established in 2010 with the aim of completing the DB with data and scores on women in the economy, and of overcoming the critique moved to the DB about its being gender-blind, and still ongoing: see wbl.worldbank.org/en/wbl (retrieved 7 December 2023)), the 'Global Indicators of Regulatory Governance' (GIRG) (released once in 2016 to explore how governments interact with the public when shaping regulations that affect their business community, and then never published again: see rulemaking.worldbank.org/en/about-us (retrieved 7 December 2023)), the 'B-Ready' project (now under preparation, with the aim of taking up the legacy of the DB "to assess the business and investment environment worldwide annually": see worldbank.org/en/businessready (retrieved 7 December 2023)).

79 See the alternative projects mentioned *supra*, at section 4.2. Global Indicators.

80 Infantino, *supra* n 56, 147. However, given that the WB discontinued the DB in 2021 following the discovery of negotiations between the WB team and individual countries regarding their scores (see worldbank.org/en/news/statement/2021/09/16/world-bank-group-to-discontinue-doing-business-report (retrieved 7 December 2023)), it is fairly possible that the scope for national influence over DB reports was much broader than had been officially acknowledged.

81 On this story, cf H. Gött, 'The ILO's Model Under Pressure: The World Bank's Employing Workers Indicator', in H. Gött (ed), *The Law of Interactions Between International Organizations. A Framework for Multi-Institutional Labour Governance* (Cham: Springer, 2020) 67–95; D. Collier and P. Benjamin, 'Measuring Labor Market Efficiency. Indicators that Fuel an Ideological War and Undermine Social Concern and Trust in the South African Regulatory Process', in Merry, Davis, Kingsbury (eds), *supra* n 17, 284–316; P. Benjamin, H. Bhorat, H. Cheadle, 'The Cost of 'Doing Business' and Labour Regulation. The Case of South Africa', *International Labour Review* 149 (2010) 73–91.

gender-blindness were the main reason underlying the establishment, in 2010, of a gender-based spin-off indicator called 'Women, Business and the Law' (WBL) reports.[82] Similarly, repeated complaints about the scant transparency of the measurement process, the fact that it was impossible to participate in it and the absence of any form of review gradually led the WB to publish all supporting materials for the reports, to enable interested parties to submit relevant information, and to subject the reports to external audits.[83]

These DB-related illustrations confirm the point stressed above in this section. Performance-based measurements exist at many different levels, that may be identified as the 'where' in which they operate. However, measurements often also subsist in complex ecologies that cut across spheres of governance, constantly interacting with forces situated outside the boundaries of their own locus of activity. The case of the DB demonstrates that the feedback and reflexive cycles created by the repetition of measurements may extend well beyond the biunivocal relationship between those who measure and those who are measured, and may embrace any person who relies on the quantitative assessments of others' performance; as a result, external reactions to a measure end up pressuring the measure itself to change.[84] Through all the linkages, interactions and feedback loops among all of the actors concerned—irrespective of where these actors are situated—quantitative measurements of performance emerge in a given setting, settle, adapt, sometimes change, and sometimes fall.[85]

5.2 *How Metrics Calculations are Performed*

We have already noted that legal metrics come in a variety of shapes and sizes:[86] scores and scorecards, rankings, ratings, benchmarking, black/white-lists, big data analysis, and hybrid quali-quantitative judgments. Moreover, performance-based measurements may be arrived at through analog or digital means (or both). We cannot enter into too much detail about the many forms that performance-based measurements can take; in what follows, we will try to highlight some features that, amidst their great variety, these measures have in common.

82 On this indicator, see n 78, as well as C. Powell, 'Gender Indicators as Global Governance: Not Your Father's World Bank', *Georgetown Journal of Gender and Law* 17 (2016) 777–807.

83 Infantino, *supra* n 56, 243–244.

84 On this dynamic, cf Zumbansen, *supra* n 23, 15–22; Tamanaha, *supra* n 23, 449–477; Berman, *supra* n 23, 2–12; Shaffer, *supra* n 23, 231–253; Cassese, D'Alterio, De Bellis, *supra* n 23, 347–355, 367–368.

85 See the authors quoted *supra* n 84.

86 See section 2. The Meaning of Legal Metrics.

Let us start from the issue of titles. Performance-based measurements often have bureaucratic names, but sometimes have nice titles or acronyms, which are supposed to make them travel smoothly in the world. The indexes named 'Freedom in the World' and 'Global Rights' are examples of this branding strategy at the global level.[87] FICO®, the formula that supports the vast majority of consumer credit assessments in the US, is the acronym of the corporation ('Fair, Isaac and Company') that devised it,[88] while the system for consumers' social credit scores managed by the Chinese company Ant Financial (a member of the Alibaba group) is evocatively dubbed 'Sesame Credit' ('芝麻信用').[89] Some Chinese cities have used catchy names for their municipal social credit systems, connecting them to the cities' rivers, flowers and other meaningful local symbols, such as the 'Qian River' ('钱江') score in Hangzhou, the 'Jasmine' ('茉莉') score in Fuzhou, the 'Little Egret' ('白鹭') score in Xiamen and the 'Seashell' ('海贝分') score in Weihai.[90]

Whether or not they have a nice title, all performance-based measurements are rooted in a methodology and a procedure, and several choices have to be made when determining them. These choices include decisions about what to count (and not to count), about the variables and dimensions that matter (and those that do not), and also about the proxies that capture information concerning these variables and dimensions (and those that do not). It is also necessary to decide on the weight assigned to each variable, dimension and proxy, the kind of data to be collected, the methods according to which non-quantitative data points should be obtained, rendered comparable with one another and transformed into numbers, and the strategies for coping with noisy, missing or incomplete data. It is particularly important to establish what information (e.g., facts, opinions, perceptions) should be gathered and in what language, where the data should be taken from (e.g., through direct measurement, self-reporting, surveys, third parties' evaluations), how data

87 On both these indicators, see section 3. Instances of Legal Metrics.

88 See fico.com (retrieved 7 December 2023); see also J.S. Hiller and L. Sain Jones, 'Who's Keeping Score?: Oversight of Changing Consumer Credit Infrastructure', *American Business Law Journal* 59 (2022) 61–121, at 71–77; Vardi, *supra* n 38, 89, 97.

89 See zmxy.com.cn/#/home (retrieved 7 December 2023). There is a substantial literature on Chinese corporate social credit systems: in addition to the authors mentioned *supra* note 38, cf A. Devereaux and L. Peng, 'Give us a little social credit: to design or to discover personal ratings in the era of Big Data', *Journal of Institutional Economy* 16 (2020) 1–19, at 5–6; Y. Chen and A.S.Y. Cheung, 'The Transparent Self under Big Data Profiling: Privacy and Chinese Legislation on the Social Credit System', *Journal of Comparative Law* 12 (2018) 356–378, at 361–363.

90 Infantino and Wang, *supra* n 38, 76.

should be processed (e.g., by humans or by machine), and who (or what) should bear specific responsibility for the above steps.[91]

As highlighted above, these elements are often revised over time, for every iteration of a measure provides an opportunity to learn from previous cycles.[92] Nevertheless, it must be underlined that all of the choices mentioned are crucial, insofar as they all influence the type and quality of results that a measure can achieve.[93] Depending on their level of complexity, they may often demand an expenditure of time and money, the hiring and training of personnel with specific skills, and/or the establishment of dedicated offices.[94] The costs of measurements also explain why methodological choices are usually dictated by practical constraints as much as (if not more than) by intentional design.[95]

It is apparent from the list of issues provided above that many methodological options are available, and there is no perfect recipe for how to build a measure.[96] Every decision comes with its own pros and cons. For instance, reliance on human judgment ensures flexibility and context-sensitivity, but is also expensive, time-consuming, and exposed to errors and subjective bias—all the more so when the measure gathers sentiments and perceptions about performance by micro- or macro-economies or individuals. One need only consider the naïveté (and blatant neglect of any basic comparative law

91 For a tentative list of the technical steps that should be required in order to design a system for measuring social performance, see D. Restrepo Amariles and J. McLachlan, 'Legal Indicators in Transnational Law Practice: A Methodological Assessment', *Jurimetrics Journal* 58 (2018) 163–209, at 183–190; D. McGrogan, 'The Problem of Causality in International Human Rights Law', *International and Comparative Law Quarterly* 65 (2016) 615–644, at 631–633; M.L. Satterthwaite and D. Kacinski, 'Quantitative methods in advocacy-oriented human rights research', in Andreassen, Sano, McInerney-Lankford (eds), *supra* n 61, 282–385, at 292–304. Needless to say, other layers of technical specialization have to be added when the quantification is partially automated, i.e., when it is largely embedded in algorithms: see, for all, Burk, *supra* n 62, 1147–1204.

92 See section 5.1. Where Metrics Can Be Found.

93 See the authors quoted *supra* n 91.

94 See 1. Introduction and section 6.2. The Dark Side.

95 It is often noted that measurements tend to count what can be easily counted, rather than counting what really matters. Among many, see Merry, *supra* n 45, 7; Davis, Kingsbury, Merry, *supra* n 17, 13.

96 There is general agreement on this point. Cf Murthy, *supra* n 36, 423–424, 443; Satterthwaite and Kacinski, *supra* n 91, 290–292; Merry, *supra* n 45, 7, 16–17; Davis, Kingsbury, Merry, *supra* n 17, 13; Broome and Quirk, *supra* n 6, 828; Cooley, *supra* n 17, 28; T. Porter, 'Making serious measures: numerical indices, peer review, and transnational actor-networks', *Journal of International Relations and Development* 15 (2012) 532–557; R. Michaels, 'Comparative Law by Numbers? Legal Origins Thesis, Doing Business Reports, and the Silence of Traditional Comparative Law', *American Journal of Comparative Law* 57 (2009) 765–795, at 786–787 (2009).

teaching) transparently displayed by one of the major producers of sentiment indicators, the United Nations Economic Commission for Europe (UNECE): "[d]ata collection and reference periods may differ among countries. Different questions may be used and questions which are intended to measure the same phenomenon may be formulated in different ways. The perception of a phenomenon and questions related to this may also vary because of cultural and other differences and may influence the comparability across countries. [...] Such differences will, eventually, affect the comparability across countries. [...] However, it should also be noted that even if individual questions differ, if the sentiment indicator is composed from a set of many questions, these differences may tend to cancel out so that the resulting indicator nevertheless may be used for international comparisons."[97]

To be sure, automated forms of measurements may be cheaper, faster and more objective. Yet the price to pay for these qualities is the enhanced rigidity and opacity of automated measures, especially when the underlying algorithms are copyrighted (as is the case for both FICO® and Sesame Credit). Moreover, the possibility that the formula underlying the measure may be secret, coupled with the fact that automated metrics are often embedded in an interface that directly implements the algorithm's suggestions and severely limits scope for double-checking and challenging its results.[98]

We do not have enough data to suggest that certain methodological patterns are more common in some sectors than in others. What we may nevertheless note is that the performance-based measures seem to position themselves along a spectrum ranging from analog, expert-driven and more qualitative forms of assessments to algorithmic (or AI-driven), mechanical and more quantitative evaluation mechanisms. For instance, performance-based assessments of justice currently seem to dominated by quantitative-qualitative, analogic forms of evaluation,[99] although the introduction of robo-judges and digitalized justice may rapidly change this.[100] At the other

97 United Nations Economic Commission for Europe (UNECE), *Guidelines on producing leading, composite and sentiment indicators* (Geneva: UNECE, 2019) 48, available at unece.org/DAM/stats/publications/2019/ECECESSTAT20192.pdf (retrieved 7 December 2023).

98 This is stressed by many. Cf, for instance, Burk, *supra* n 62, 1186–1187; O'Neil, *supra* n 1, 141–160; Pasquale, *supra* n 39; D. Keats Citron and F. Pasquale, 'The Scored Society: Due Process for Automated Predictions', *Washington Law Review* 89 (2014) 1–33, at 14; Citron, *supra* n 64, 1275.

99 See the authors quoted *supra* n 25.

100 N. Wang and M.Y. Tian, "Intelligent Justice': AI Implementations in China's Legal Systems', in A. Hanemaayer (ed), *Artificial Intelligence and Its Discontents. Critiques from the Social Sciences and Humanities* (Cham: Palgrave MacMillan, 2022) 197–222; Sourdin, *supra* n 63.

extreme, the digital domain is dramatically dominated by purely quantitative, machine-led tools.[101] Measures in other sectors lie somewhere in between these two poles, mixing qualitative, expert-driven judgments with automated, quantitative mechanisms.

5.3 *Whose Metrics?*

The analysis carried out so far has shown that the dynamic initiated and supported by performance-based measurement involves a plethora of actors that quantify processes, activities and results attributable to an equally rich array of other actors. Furthermore, these dynamics often extend beyond the relationship between those who set up the measurement and those who are subject to it, involving wider networks of stakeholders. These stakeholders are sometimes interested in participating in the construction and management of the measure; at other times they may criticize and resist the quantification exercise; and at other times, they may embrace the initiative by imitating it or relying upon it for their own purposes. Be it as it may, the complexity of the network of actors surrounding performance-based measurements makes the answer to the question 'by whom and for whom are legal metrics made?' less obvious than it might appear to be at first sight.

The examples of performance-based measures surveyed above demonstrate that this question requires at least a bifurcated answer. When considered in the light of who measures, and for the benefit of whom, a neat (at least, in theory) distinction emerges between initiatives aiming to quantify one's own performance and initiatives quantifying the performance of others (although in practice the dividing line between the two types of metric is often unclear).

In a typical scenario, someone—a state, an agency or a company—measures its own performance or the performance of its own personnel. This is the case for measurements carried out by state authorities in the field of justice, legal research, education, healthcare and public services, as well as for human rights self-assessments prepared by states and corporations.[102] Although these quantitative exercises may also send out a signal to interested stakeholders or to the general public, their primary aim seems to be self-improvement and self-regulation. In other words, the entity producing the measurement is the same as both the measured entity and the beneficiary of the measurement. Consequently, initiatives of this kind raise few problems in terms of legitimacy, as the power of self-measurement is included within the power of self-regulation.

101 See the authors mentioned *supra* n 62–64.

102 See the illustrations provided in section 3. Instances of Legal Metrics.

The structure of the relationship between the actors affected by performance-based measurement becomes notably different in other cases. Consider, for instance, universities, journals and researchers that are scored by specialized companies,[103] global indicators comparing the legal architecture of the world's countries,[104] and the algorithmic profiling of businesses' and individuals' online behavior.[105] Authors of these measurements notably differ from those who are subject to them: the rationale for quantification here is no longer self-improvement and self-regulation, but rather the 'improvement' and regulation of others. In the vast majority of these scenarios, performance-based measurements are carried out by actors that have no connection with, and little to no authority over, those who are measured. In a limited number of instances, the measurement may still derive some legitimacy from the specific expertise of its author, or from its position vis-à-vis those who are measured (such as when the World Bank measures countries' business-friendliness or a food delivery platform measures the punctuality of its riders). However, in other cases (for instance the many NGO-led global indicators or companies' tracking of online activities[106]), those who carry out the assessment have no expertise or status justifying their intervention. On the contrary, it is the very act of measuring that provides them with the experience and position to measure, thereby legitimizing ex post facto their power to quantitatively assess the performance of others.[107] It should be added that many of these measurements, especially those with a transnational or global scope, share the distinctive feature of being authored by actors based in the West (and, more often than not, in the US), who often claim the authority to assess compliance with allegedly universal standards by everyone else.[108] These measurements may in some way be useful for those

103 See the cases mentioned *supra* n 25, 27–28 and 75.

104 See sections 3. Instances of Legal Metrics and 5.1. Where Metrics Can Be Found.

105 See the illustrations provided in section 3. Instances of Legal Metrics.

106 See section 3. Instances of Legal Metrics.

107 This paradoxical *ex post facto* legitimization has been noted by many. See Catá Backer and McQuilla, *supra* n 37, 2; Broome and Quirk, *supra* n 6, 823; Cooley, *supra* n 17, 24; Espeland and Sauder, *supra* n 25, 36. Well-known examples of this effect are the role and authority obtained in global governance by transnational NGOs measuring states' respective performances: for the examples of Freedom House and Transparency International, see respectively *supra* n 33 and 35.

108 Among the many who have noted this, cf S.S. Bush, 'The Politics of Rating Freedom: Ideological Affinity, Private Authority, and the Freedom in the World Ratings', *Perspectives on Politics* 15 (2017) 711–731, at 722–724; J. Kroncke, 'Law and Development as Anti-Comparative Law', *Vanderbilt Journal of Transnational Law* 45 (2012) 477–555, and in particular 479 and 538; L. Clegg, 'Our Dream is a World Full of Poverty Indicators: The US, the World Bank, and the Power of Numbers', *New Political Economy* 15 (2010) 473–492, at 474, 481; W. Twining, *Globalisation and Legal Theory* (London: Butterworths, 2000) 158; Strathern, *supra* n 4, 2.

who are measured and for interested stakeholders. However, it is certain that they also help their authors to affirm and reinforce their own authority. We can go so far as to assert that those in power often measure as much as those who measure come to power.

5.4 *Why a Metrics Approach is Adopted*

As noted above, if one leaves aside sector-specific rationales (such as better justice or more personalized services for online consumers), it seems at first glance that the most commonly recurring theme for measuring both one's own and others' performance is a somewhat undefined wish to promote (self-) improvement. The notion of improvement is of course a tricky one, insofar as it evokes the idea of progress and betterment, as if there were a universal consensus on the answer to questions about what improvement should be achieved, for whom, at what cost, and what for.[109] Yet, as vague as it may be, the ideology of improvement sustains the shared management-oriented belief that measuring processes and results unleashes efficiency gains, because, according to that frequently repeated catchphrase, 'what gets measured gets done'.[110]

This belief in the improvement capacity of measures is rooted in well-known features of quantitative performance assessments. Metrics are supposed to provide objective and impartial knowledge of social phenomena that can help carry out diagnostic analysis, highlight problems that need to be fixed, identify patterns for corrective action and reward appropriate behavior. Furthermore, by repeatedly monitoring progress, by naming and shaming bad performers and by singling out good performers, metrics aim to foster change (either directly or through pressure exerted by relevant stakeholders) in the behavior of those who are measured, stimulating competition among them, virtuously orienting their actions, and in any case making them accountable for how they perform.[111]

---

109 This is not to mention the fact that the notion of improvement carries itself the burden of ideology and history: see, for all, P. Slack, *The Invention of Improvement: Information and Material Progress in Seventeenth-Century England* (Oxford: OUP, 2014) esp 1–14.

110 'What gets measured gets done' is for instance the title of one of the sections of final reports by the United Nations on the MDGs: United Nations, *The Millennium Development Goals Report 2015* (New York: United Nations, 2015) 10, at un.org/millenniumgoals/2015_MDG_Report/pdf/MDG%202015%20rev%20(July%201).pdf (retrieved 7 December 2023).

111 See, in addition to the authors cited *supra* n 56 and 60–64, Infantino, *supra* n 17, 356; K.E. Davis, B. Kingsbury, S. Engle Merry, 'Introduction: Global Governance by Indicators', in Davis, Fisher, Kingsbury, Merry (eds), *supra* n 60, 3–28.

Despite all of these features, the capacity of performance-based measurements to stimulate improvement is much less straightforward than it might seem. At least three problematic aspects help to explain why this is the case (and also why the myth of improvement through measurement is constantly retold in spite of the contradictory evidence supporting it).

A first problem arises with regard to the very definition of (good/bad performance and therefore) improvement. As noted a few lines above, the notion of improvement implies an ideal to be pursued. Especially in cases in which those who measure aim to nudge others into improvement (even though, as noted above,[112] they often have little pre-existing expertise or authority in the field), that ideal may not be fully agreed upon or shared by all actors involved. For instance, according to tech companies and data vendors, getting people to stay connected for longer and to spend more time engaging online while data about their performance are constantly tracked pursues the aim of providing consumers with better services and offers that are more 'personalized' (i.e. mass-customized on the basis of algorithmic correlations) in line with their needs. However, online tracking also furthers other goals—such as obtaining better knowledge from consumer data in order to exploit vulnerabilities, create unwanted needs, setting prices as high as consumers profiled are willing to pay, and also predict and manipulate desires and opinions—not all of which could be defined as an improvement from the consumers' perspective.[113] In other cases, the ideal against which performance is measured may be tainted by bias or ideology. The DB again provides a good illustration of this point. The DB reports measured countries' performance in terms of the business friendliness of their legal architecture on the basis of a myriad of assumptions. Assumptions included the idea that the common law tradition is more conducive to economic growth than the civil law one, that all legal systems can be viewed as variations of either common law or civil law, that rules can be easily transplanted from one place to the other, that less regulation is always better, and that there is one 'right' answer to any business's

112 See section 5.2. How Metrics Calculations are Performed.

113 C. Riefa, 'Protecting Vulnerable Consumers in the Digital Single Market', *European Business Law Review* 33 (2022) 607–634; L.E. Willis, 'Deception by Design', *Harvard Journal of Law and Technology* 34 (2020) 116–190; Burk, *supra* n 62; Couldry and Mejias, *supra* n 6, 83–112; Zuboff, *supra* n 28; Lupton, *supra* n 28, 64–87.

114 See, among many, Broome, Homolar, Kranke, *supra* n 48; Garoupa, Gómez Ligüerre, Mélon, *supra* n 31; Michaels, *supra* n 96, 786–787; B. Fauvarque-Cosson and A.-J. Kerhuel, 'Is Law an Economic Contest? French Reactions to the Doing Business World Bank Reports and Economic Analysis of the Law', *American Journal of Comparative Law* 57 (2009) 811–829, at 814–815.

legal problem, which invariably coincides with the common law solution.[114] These assumptions were clearly in line with the WB's promotion of neo-liberal globalization; however they were also all debatable, if not plainly wrong.[115] Unsurprisingly, many noted that countries' improved performance in the DB report did not reflect improved performance in the real world—not even according to the limited meaning of 'improvement' embraced by the WB.[116]

This leads us to a second, structural problem in using performance-based measurements as tools for improvement. No matter what notion of improvement is embraced, quantitative assessments invite those who are subject to them to focus their efforts on the metric being counted so as to produce better results. However, whatever is being counted always captures only a fraction (often, the most easily countable fraction) of what the measured actor is doing. As a result, the act of counting often focuses the attention of those who are measured on a few variables, while interest is lost in any dimension that, although it may be crucial, is not being counted. By way of illustration, a legal research measure that only rewards English-language monographs and articles in indexed journals drives researchers to write in English and to submit papers to these journals, thereby underestimating the quantity and quality of research published in a different format or language.[117] The example also reveals how, by changing the behavior of those who measured, the act of measurement easily generates self-fulfilling prophecies: the adhesion by (some of) the measured actor to the chosen variable confirms ex post, and reinforces, the formal validity of the variable, irrespective of its actual meaningfulness.[118] It is clear from this that measures often tell us little about the dynamics underlying the process that is measured. Rather, what measures track (and stimulate) is improvement in *measured* performance. Performance-based measures monitor the measured actor with regard to chosen variables, rewarding good performers with higher results. However, they are unable to appreciate whether these efforts came at a cost, whether unmeasured variables worsened, and whether overall performance is (not only nominally better, but also) actually better—or even simply good.[119]

115 See the authors *supra* n 114 and 116.

116 Cf McCormack, *supra* n 31, 649–676; A. Perry-Kessaris, 'The re-co-construction of legitimacy of/through the Doing Business indicators', *International Journal of Law in Context* 13 (2017) 498–511; B. Arruñada, 'How Doing Business Jeopardizes Institutional Reform', *European Business Organization Law Review* 10 (2009) 555–574; B. Arruñada, 'Pitfalls to Avoid When Measuring the Institutional Environment: Is "Doing Business" Damaging Business?', *Journal of Comparative Economics* 35 (2007) 729–747.

117 A. Jakubowski, 'Quantification and Parameterization of Legal Research: The Case of Poland', in Bussani, Cassese, Infantino (eds), *supra* n 25.

118 On all points, see Espeland and Sauder, *supra* n 25, 11–12.

119 See, on all points, Espeland and Sauder, *supra* n 25, 11–16.

A third additional reason why performance-based measures may not improve performance (other than in relation to the measure itself) stems from the unintended effects of social quantification. As noted above, measurements tend to focus the attention of the measured actor on what is counted, and to divert it from variables and aspects that are not counted. It is well known that this mechanism generates a number of side-effects (besides self-fulfilling prophecies). These are not limited to the rather rare cases in which insistence on the measurement of certain variables exhausts the measured actor and triggers resistance, including—where this choice is available—disengagement from or active manipulation of the measure (a reaction often documented with self-tracking apps).[120] There is a much more common unintended effect of performance-based measures: quantification stimulates the adoption by those who are measured of gaming strategies and rank-seeking tactics—i.e. a behavior that formally conforms to the ideal embraced by the measure but that implies little or no change in practice.[121] After all, once a measure is in place, what the measured person often wants is to achieve better scores or, in other words, improvement certified by the measurement, as our initial illustration with students' grades reminds us.[122] Therefore, as any student knows, cheating is a viable option. Whenever one can obtain better scores through minimal or symbolic effort, or through the manipulation of apparent data and numbers, gaming the measure may turn out to be the most efficient and cost-effective approach for dealing with it, as such a choice implies no change and no cost other than those required by the act of cheating itself. Needless to say, gaming strategies and rank-seeking tactics pollute measures, and increase the possibility that the 'improvement' captured by quantitative assessments will not be reflected by any improvement at all in the real world.

All the above takes us back to our initial question. Given that performance-based measures often fail to improve performance, why are they flourishing? In order to answer the question it would be necessary to reach beyond the mantra of measurements that result in efficiency gains and look for other rationales justifying the all-encompassing turn to social quantification. What we can do is point to some of the most frequently recurring patterns for establishing measures of one's own or others' performance.

120 M. Tanninen, 'Contested Technology: Social Scientific Perspectives of Behaviour-Based Insurance', *Big Data and Society* (2020) 1–14, at 8; H. Jeanningros and L. McFall, 'The Value of Sharing: Branding and Behaviour in a Life and Health Insurance Company', *Big Data and Society* (2020) 1–15, at 9–11; Couldry and Mejias, *supra* n 6, 193–194.

121 Espeland and Sauder, *supra* n 25, 29–33.

122 See section 1. Introduction.

Perhaps the clearest pattern of all is the following: measuring performance often allows someone to control what would otherwise be beyond anybody's control. For example, in the West it is (in principle) not allowed to encroach upon judges' independence or researchers' scientific freedom, but it is permitted to measure their performance. Forcing states to adhere to standards that are only partially enshrined in international treaties (such as those on corruption, the rule of law, and openness to foreign investment) is not possible, yet the measurement of their performance against these standards is. Requiring multinational companies to abide by ESG/CSR obligations is impossible in the absence of any domestic or international obligation to do so, but involving them in (self-)measuring exercises is feasible and laudable. Manipulating people's desires is not commendable, while measuring online users' preferences and using this data to provide them with more 'personalized' services and offers is. From this point of view, measuring performance allows control (some would say: surveillance) to creep into areas in which regulation is either forbidden or unfeasible, while also reinforcing (or, as often happens in cases involving the measurement of others' performance, actually creating) the authority of those who make the measure vis-à-vis those who are measured.[123]

Beyond controlling what would otherwise be unchecked, many other rationales underly the spread of performance-based measurement. Especially when done in relation to one's own performance, measurement often offers a way of proving or advertising to the public one's more or less serious commitment to improvement.[124] Sometimes, especially when the performance of others is being measured, measurement is a commercial strategy that enables the actor making the measurement to exist or survive in a crowded and competitive marketplace.[125] On other occasions—no matter whose

123 This is why many commentators emphasize that measurements are a 'technology of governance': see, among many, Kelley and Simmons (eds), *supra* n 47; N. Bhuta, D.V. Malito, G. Umbach, 'Introduction: Of Numbers and Narratives—Indicators in Global Governance and the Rise of a Reflexive Indicator Culture', in Malito, Umbach, Bhuta (eds), *supra* n 60, 1–29; Broome and Quirk, *supra* n 6; Cooley, *supra* n 17; Merry, *supra* n 45.

124 One need only consider self-assessments by corporations of their respect for ESG/CSR standards, which are often carried out in order to showcase companies' efforts and to capture the attention of ethically-minded consumers, while avoiding any real obligation: see Auld and Gulbrandsen, *supra* n 37, 398.

125 See Cooley, *supra* n 17, 21–22 ("An NGO or IO might also use rankings and ratings to 'flag-pant' or to brand themselves as the pivotal organizations advocating a particular cause or global concern. […] Branding is especially necessary in what is now a crowded market for advocacy organizations").

performance are measured—some measures are created out of other pre-existing measures, as a way for the former to imitate or counter the latter.[126] And sometimes, quite simply, one measures because one can. The dataism affecting our contemporary societies actually invites the qualitative aspects of human experience to be converted into quantitative (and possibly computer-ready) information whenever possible.[127]

5.5 *What Effects Metrics Trigger*

We saw in the two previous sections that performance improvement is an oft-repeated mantra for establishing a social measurement system, and yet is not often—if ever—what social measurements actually do in practice. What are the practical effects produced by metrics? We will deal in this section with the most mundane consequences of performance-based measures, leaving to the next section an analysis of the broader impact of social quantification as a regulatory technique.

A preliminary caveat must be made. Tracing the effects of performance-based measurements is extremely hard, because in order to do so it would be necessary to demonstrate patterns of causality between the measure and some real-word change; however, these are often impossible to prove. There are several reasons for this.[128] Much of the effects of quantitative measures often remain invisible, involving imperceptible shifts in the identity and power of those who produce and use them, in the vocabulary of policy-making, as well as in the way in which problems are defined and responses are elicited. No matter how important these shifts are, their occurrence may be unconscious, unverbalized, undisclosed—and therefore hard to document. The opposite may hold true as well: lip-service may be paid to measurements when justifying some change, even if these measurements played practically no role in that change. The adoption by the measured actors of gaming strategies and rank-seeking tactics additionally complicates the picture, obliging to distinguish between *measured* change and *real-word* change. In addition, even when it is possible to detect some real-world modification, it is still difficult to

126 For some illustrations of the spin-off projects and the alternative initiatives that the DB has generated, see section 5.1. Where Metrics Can Be Found.

127 The meaning of dataism is not unequivocal, but is generally associated with the idea that data (is the only thing that) matters: Y.N. Harari, *Homo Deus. A Brief History of Tomorrow* 428–462 (New York: HarperCollins, 2015); S. Lohr, *Data-ism: The Revolution Transforming Decising Making, Consumer Behavior, and Almost Everything Else* (New York: HarperCollins, 2015).

unequivocally ascribe the observed modification to a measure, because change is always determined by a multiplicity of factors, of which performance-based measurements are but one. Last but not the least, investigating all of the above would require working with methodological tools that are beyond lawyers' (and surely our own) expertise.

Be it as it may, the available literature provides us with some information about the real-world consequences of performance-based measurements. Let us try to summarize the results, starting from the effects of measurements on those who are subject to them and then looking at their effects on other actors.

As far as those who are measured are concerned, sometimes performance-based measurements clearly come with strings attached. Judges whose activity is not well evaluated may have their careers stall in Brazil, China, India, Mexico and Switzerland.[129] A law school whose members carry out (what the measure assesses as) outstanding research in Poland may secure special funding, and conversely faculties whose members do not perform well may be denied funding.[130] Chinese social credit initiatives reward good performers and deny or heighten the costs of access to services for bad performers.[131] Most often, however, measures do not trigger any official consequence: consider for instance the quantitative assessments of multinational companies' compliance with ESG/CSR standards, and the WB's measurement of states' performance in promoting foreign investment.[132] Nevertheless, the absence of foreseen consequences does not imply that such consequences do not exist. As has repeatedly been highlighted throughout this article,[133] by combining apparently objective descriptions with evidence-based prescriptive suggestions, measures easily become performative, pressuring the measured actor to devise strategies

128 On these problems, see Infantino, *supra* n 17, 356–360.

129 See the authors *supra* n 25.

130 Jakubowski, *supra* n 117.

131 See the authors *supra* n 38 and 89. Other examples include banks advising prospective or actual clients that their creditworthiness will be measured by FICO® or equivalent scores, or platforms, search engines and apps informing users through their terms and conditions about the collection, processing and uses of data relating to their behavior (subject to the special characteristic that, in such cases, users' acceptance of data collection is often the but-for condition, and in actual fact the consideration, for the service received). See, on all these points, Couldry and Mejias, *supra* n 6, 101–105; O'Neil, *supra* n 1, 141–160; Pasquale, *supra* n 39, esp. 1–14.

132 See the illustrations *supra* n 31–32 and 37.

133 See sections 5.1. Where Metrics Can Be Found–5.3. Whose Metrics?.

to cope with them, ranging from compliance attempts to gaming techniques to resistance tactics.[134]

The impact of performance-based measurements on actors other than those who are measured is quite a different issue. The uses (and effects) of quantitative measures by third parties may be intended or not: very often, those who measure their own performance conceive of the initiative as a purely internal procedure, while those who measure others' performance anticipate an interest in the measurement on the part of some stakeholders. This is particularly clear where quantitative assessments are only available upon subscription (as is common in the field of academic rankings, ratings, and personal credit scoring[135]), as well as in cases involving automated online profiling, since the transfer and re-use of data lie at the core of the digital economy.[136] However, some instances of measurements of others' performance explicitly forbid or caution against the re-use by third parties of the results obtained, as often occurs in relation to global measurements of states' performance.[137] Nonetheless, no matter what the position of the measurement producer is, uses by (and effects on) third parties of measures are frequent, and usually play a fundamental role in solidifying the success and legitimacy of the measures themselves.

This is because, as simplified, decontextualized and easy-to-use information, measurements can travel long distances and end up in unforeseeable places. After they have been produced, quantitative assessments often take on a life of their own: they may be appropriated and used by anybody, for purposes that may be more or less distant from those that originally inspired their creation. Aside from the example of school grades provided above,[138] we need only consider how the numbers provided by corporate rankings of journals, articles and individual academics are relied upon by (researchers themselves, universities, publishers, and) domestic authorities in their own

134 The results of these strategies may sometimes be tragic. Botero, Pinzon-Rondon, Pratt, *supra* n 25, 54–58 report that a quantitative incentive established in 2002 for policemen fighting against the Fuerzas Armadas Revolucionarias de Colombia (FARC)—and directly linking salary increases to the number of guerrilla fighters killed in action—resulted in the mass killings of innocent civilians whose corpses, dressed in FARC uniforms, were then claimed to be those of FARC combatants; the practice is known as the 'fake positive' scandal, for which Colombia was investigated before the International Criminal Court: see icc-cpi.int/colombia (retrieved 7 December 2023).

135 See the illustrations *supra* n 27, 30, 38.

136 See *supra* n 39, 62–64.

137 For some examples, see Infantino, *supra* n 56, 215.

138 See section 1. Introduction.

evaluations of the performance of higher education systems.[139] We already saw that the DB inspired several spin-off projects and even some concurrent initiatives;[140] what we should add now is that the scores produced each year by the DB report were also used by WB offices for assessing countries' eligibility for loans, identifying areas for reform and verifying whether countries had implemented the conditions attached to loans.[141] Consider also how sovereign credit ratings are relied on by governments, central banks, investors, and voters on a daily basis to assess the trustworthiness of sovereign bond issuers and, more generally, the solidity of a given country's economy.[142] It is well-known, especially in common law countries, that a plethora of actors over and above banks (including employers, insurance companies, and landlords) resort to FICO® and similar credit scores for evaluating the overall creditworthiness of their prospective counterparties.[143] Similarly, where algorithms are used to measure online performance, it is well documented that these data are repeatedly re-sold, re-used and aggregated to help predict people's behavior (further heightening the risk of social conservativism, the perpetuation of historical and systemic bias, and social exclusion of the most vulnerable and disadvantaged).[144] Sometimes uses of measurements by third parties may be counter-intuitive. One anecdote concerning an episode experienced directly by a Colombian colleague when she was conducting a field study for an NGO[145] is telling in this regard. While interviewing the NGO staff, she suddenly heard an employee shout 'Hooray! We are a failed State!'. The employee was celebrating the fact that Colombia qualified (again) as a failed state according to the 'Fragile States Index' published by the American NGO Fund for Peace.[146]

139 See the authors quoted *supra* n 25.

140 See section 5.1. Where Metrics Can Be Found.

141 T. Krever, 'Quantifying law. Legal indicator projects and the reproduction of neoliberal common sense' (2013) 34 *Third World Quarterly* 131–150, at 145. The scores of the WBL, one of the offshoots of the DB, are currently included in the formula used by the US development agency Millennium Challenge Corporation in order to distribute its funds: Powell, *supra* n 82, 779.

142 See the authors quoted *supra* n 30.

143 Burk, *supra* n 62, 1163–1166; O'Neil, *supra* n 1, 141–160; Pasquale, *supra* n 39, 24–26; Citron and Pasquale, *supra* n 98.

144 Among many, cf Burk, *supra* n 62, 1163–1166, 1181–1186; Zuboff, *supra* n 28; Umoja Noble, *supra* n 39; Eubanks, *supra* n 38; S. Barocas and A.D. Selbst, 'Big Data's Disparate Impact', *California Law Review* 104 (2016) 671–732; O'Neil, *supra* n 1; Citron and Pasquale, *supra* n 98, 1–33.

145 L.F. Buchely Ibarra, 'Indicators as a Form of Resistance', *International Law Journal* (*Universidad Javeriana*) 25 (2014) 225–265, at 226.

146 See fragilestatesindex.org (retrieved 7 December 2023).

What was the reason for celebration? The bad result ensured that the NGO could continue to seek international funding.

More research, possibly combining legal expertise with other methodological lenses, is needed to trace the consequences of performance-based measurements, both for the measured actors and for other stakeholders. Our findings nevertheless enable us to highlight that measures often change the set of expectations and incentives for the actors concerned, producing effects that in many cases are comparable to, and even more effective than, those associated with regulation. The next section will investigate exactly how and to what extent performance-based measures relate to the law.

6 Legal Metrics as a Regulatory Technique

It is apparent from the previous sections that social quantification in the form of performance-based measurement is widespread and operates as a technology of distance.[147] It has also been seen that, once measures exist, they often trigger (intended and unintended) behavioral changes in those who are measured and other relevant stakeholders, thus modifying the world in which they operate. Due to this tendency to become performative and to change (while also being changed by) the reality they apply to, social quantification studies have long noted that performance-based measures may easily turn into tools for governance and regulation, controlling and spreading standards for behavior in a way akin to (and often more effective than) what traditionally is conceived of as law.[148]

It is well-known that the notions of 'governance' and 'regulation' are often used to mean quite different things,[149] and that people, including lawyers, tend to disagree about what they think the 'law' is.[150] It is equally well known

147 See the authors *supra* n 6.

148 In addition to the authors *supra* n 123, see Borges Fortes, Baquero, Restrepo Amariles, *supra* n 64; L. Ulbricht and K. Yeung, 'Algorithmic Regulation: A Maturing Concept for Investigating Regulation of and through Algorithms', *Regulation and Governance* 16 (2022) 3–22; M.K. Land, 'The Problem of Platform Law: Pluralistic Legal Ordering on Social Media', in Berman (ed), *supra* n 23, 975–994; Broome, Homolar, Kranke, *supra* n 48; M.M. Siems and D. Nelken, 'Global social indicators and the concept of legitimacy', *International Journal of Law in Context* 13 (2017) 436–449; Davis, Kingsbury, Merry, *supra* n 111; Hildebrandt, *supra* n 64, 11.

149 Braithwaite, Coglianese, Levi-Faur, *supra* n 11, 3.

150 For a reasoned summary of the most accepted theories in this area, see M. Croce, *Self-sufficiency of Law. A Critical-institutional Theory of Social Order* (Cham: Springer, 2012);

that 'regulation' refers to a large array of elements, including politics, policies, (judicial, quasi-judicial, extra-judicial), 'nudges',[151] 'co-design',[152] institutions, compliance, and formal and informal controls.[153] This is why regulation is itself a product of different framings and power dynamics. From our perspective, however, what matters is that the commonly envisioned goal of regulation as being beneficial for the public interest can by no means be taken for granted, also because it is hard to maintain that all actors operating in a competing business, social, political environment share the same values, needs, and interests.[154] Within this context, performance-based measures have the ability to project their (often unarticulated) ideal views onto the measured actors as well as onto other stakeholders, and to stimulate from them reactions that reinforce the validity and cogency of these ideals. Insofar as they embody principles and values that are perceived of by the relevant community as binding, performance-based measures may well be said to have normative implications. Quantitative assessments may even be regarded as more powerful tools than regulation, insofar they simultaneously articulate a standard for behavior and apply it, rewarding compliance and sanctioning breaches.[155]

This conclusion is shared by the limited legal research into measures in the social world surveyed above.[156] For instance, the legal literature on indicators often notes that indicators "hold an intrinsic normative quality",[157] "function as

see also Zumbansen, *supra* n 23, 15–22; Berman, *supra* n 23, 2–12; Bussani, *supra* n 23, 3127–3160; Shaffer, *supra* n 23; Cassese, D'Alterio, De Bellis, *supra* n 23, 347–355, 367–368.

151 A seminal study is that by Thaler and Sunstein, *supra* n 54. Within the endless literature on nudging, those contributions that are of some interest for our present purposes include C. Einfeld and E. Blomkamp, 'Nudge and co-design: complementary or contradictory approaches to policy innovation?', *Policy Studies* 43 (2021) 901–919; S. Cassese, 'Exploring the Legitimacy of Nudging', in A. Kemmerer, C. Möllers, M. Steinbeis, G. Wagner (eds), *Choice Architectures in Democracies: Exploring the Legitimacy of Nudging* (Oxford: Bloomsbury, 2016) 241–246; M. Whitehead, R. Jones, R. Howell, R. Lilley, J. Pykett, *Nudging All Over the World: Assessing the Global Impact of the Behavioural Sciences on Public Policy* (Aberystwyth: Economic and Social Research Council, 2014).

152 See, e.g., Einfeld and Blomkamp, *supra* n 151; L. Kimbell, *Applying Design Approaches to Policy Making: Discovering Policy Lab* (Brighton: University of Brighton, 2015).

153 See, generally, Levi-Faur, *supra* n 11, 3 ff.

154 C. Papaevangelou, 'The existential stakes of platform governance and online content regulation: a critical conceptual model', *Open Research Europe* 1:31 (2021).

155 Davis, Kingsbury, Merry, *supra* n 17, 18–19; this ability is particularly strengthened whenever the measurement is performed by an algorithm: see Burk, *supra* n 62, as well as the authors quoted *supra* n 171.

156 See sections 4.2. Global Indicators–4.3. Algorithms and the Law.

157 D.V. Malito, N. Bhuta, G. Umbach, 'Conclusions: Knowing and Governing', in Malito, Umbach, Bhuta (eds), *supra* n 60, 503–512, at 507.

implicit or even explicit standards",[158] operate as "regulatory devices",[159] are "comparable to law"[160] or even constitute "soft law"[161] themselves. The same kind of remarks are nowadays commonly made with regard to algorithmic and AI-driven quantification. The revolutionary intuition by Lawrence Lessig that code acts as law insofar as it provides a socio-technical infrastructure that determines what people can do and shapes their mutual expectations for interaction[162] is now routinely conveyed by the idea that code produces "algo norms",[163] "algorithmic law"[164] and "algorithmic regulation".[165] We therefore stand on the shoulders of giants when we claim that performance-based measurements may qualify as "new forms of regulation [that] are fit subjects for legal research and teaching whether or not they are conceived of as 'law'".[166]

But what are the defining features of this regulatory technique? We argue that, as a regulatory technique, performance-based measures present many distinctive traits, the strengths and weaknesses of which deserve to be highlighted separately.

158 Krever, *supra* n 94, 132; along the same lines, see also Davis, Kingsbury, Merry, *supra* n 111, 15; Davis, Kingsbury, Merry, *supra* n 17, 18–19; B. Frydman, 'Prendre les standards et les indicateurs au sérieux', in Frydman and Van Waeyenberge (eds), *supra* n 60, 5–68, at 21–25.

159 S. Cassese and L. Casini, *supra* n 60, 465–474, at 466; see also B. Frydman and W. Twining, 'Preface. A symposium on global law, legal pluralism and legal indicators', *Journal of Legal Pluralism and Unofficial Law* 57 (2015) 1–5, at 5.

160 Davis, Kingsbury, Merry, *supra* n 17, 2; see also J.G. Kelley and B.A. Simmons, 'Politics by Number: Indicators as Social Pressure in International Relations', *American Journal of Political Science* 59 (2015) 55–70 (noting that indicators "have effects akin to regulation").

161 Murthy, *supra* n 36, 390; see also M. Bussani, 'Deglobalizing Rule of Law and Democracy: Hunting Down Rhetoric Through Comparative Law', *American Journal of Comparative Law* 67 (2019) 701–744, esp. at 718–720; Restrepo Amariles, *supra* n 60, 17; Merry, *supra* n 45, 11.

162 Lessig, *supra* n 64, esp. 77–84.

163 H. Hydén, 'AI, Norms, Big Data, and the Law', *Asian Journal of Law and Society* 7 (2020) 409–436.

164 Catá Backer and McQuilla, *supra* n 37, 1; see also Hildebrandt, *supra* n 64, 11–12.

165 Borges Fortes, Baquero, Restrepo Amariles, *supra* n 64; Ulbricht and Yeung, *supra* n 148; Yeung and Lodge (eds), *supra* n 63; M. Hildebrandt, 'Algorithmic Regulation and the Rule of Law', *Philosophical Transactions of the Royal Society Series A: Mathematical, Physical, and Engineering Sciences* 276 (2018) 20170355; J. Grimmelmann, 'Regulation by Software', *Yale Law Journal* 114 (2005) 1719–1742.

166 Frydman and Twining, *supra* n 159, 5.

6.1 *The Bright Side*

Starting with the strengths, there are at least three advantages of relying on performance-based measures.

One first specificity of such measures relates to the way in which measurements act on the world. As already recalled,[167] performance-based measurements have the rare ability to embody a standard of behavior, to provide carrot-and-stick incentives to comply with that standard, and to verify the level of compliance, all at the same time (and all the more so where the measurement is automated).

A second strength of performance-based measurements arises from the fact that they are usually conceived of as technical tools providing neutral, empirical and evidence-based knowledge about the social phenomena they apply to.[168] The allegedly technical nature of quantification initiatives implies that the solutions they offer are often eagerly accepted by their addressees and relevant stakeholders as a reasonable basis for policy-making, decisions and actions.

Precisely because they are conceived of as technical exercises, a third strength of performance-based measures lies in their political agility (or invisibility, if you will). The establishment of a quantitative assessment does not typically require any political decision or diplomatic bargaining, and its management is not constrained by requirements of good regulatory governance such as transparency and participation: "the traditional contentions inherent in the formation of policy choices and the legal structures that support them [are] transferred to the development of data, its analytics, and the algorithms that are meant to reflect judgment".[169] Simply put, the alleged technicality of quantification helps to insulate it from political scrutiny and legal challenges.[170] Whenever quantification is automated, the significance of all these features—the combination of regulation-control-enforcement, the rapidity, technicality, invisibility and unchallengeability of the measuring process—increases exponentially.[171]

---

167 See the authors quoted *supra* n 155.

168 See especially the authors *supra* n 56 and 60–62.

169 Catá Backer, *supra* n 38, 166. Political invisibility may explain why, as already noted (see *supra* n 155), reliance on performance-based measurements is widespread in contexts that are not disciplined by official law.

170 On all these features, cf Infantino, *supra* n 17, 356; Davis, Kingsbury, Merry, *supra* n 111.

171 Among others, see Burk, *supra* n 62; A.D. Selbst and S. Barocas, 'The Intuitive Appeal of Explainable Machines', *Fordham Law Review* 87 (2018) 1087–1139, at 1092–1096 (2018); Barocas and Selbst, n 144, 671–731, at 677–694; Hildebrandt, *supra* n 64, 11–12; Citron and Pasquale, *supra* n 98, 5–6, 11–16; Grimmelmann, *supra* n 165, 1729–1730.

6.2 *The Dark Side*

As the three points of strength listed above suggest, performance-based measurements also have darker sides. Let us start by noting that, as is also emphasized by a plethora of research on social quantification,[172] measures convey objective and empirical knowledge on what they measure only inasmuch as they can capture comprehensive information concerning the measured phenomena. More often than not, measures are not able to do so: as fast, top-down and routinized forms of control, performance-based measurements rely on standardized and simplified variables (e.g., the number of judgments rendered in a given time span, the impact factor of the journal in which a scientific article is published, web users' online browsing speeds) that require context and complexity to be erased in order to obtain comparable results.[173] Anything that does not fit in with the standardized categories is either forced into them or discarded. In this regard as well, the more quantitative and automated the assessment is, the more inflexible it tends to be, and the more information gets lost in the process.

What is lost in the process is notably a taste for quality: when measuring performance, it does not matter whether the outcome of a judicial opinion was fair, how good a scientific article is or the purpose of web browsing; at best, these aspects may be taken into account indirectly (for instance, by proxies adding to the picture whether the judgment was reversed, how many citations the article received, and which websites were visited). Measuring, in other words, shifts the attention from the substance of what is done to the form of what can be counted as (having been correctly) done.

This shift from substance to form implies a heightened significance of evidence. Doing something bears little relevance for a measure if there is no proof of what has been done. Usually, proof should be given using the standardized forms of evidence (e.g., self-reporting, third-party audits, log-ins,

172 In addition to the seminal works by Desrosières, *supra* n 6, and Porter, *supra* n 6, cf de Souza, *supra* n 61, 27–50, 112–173; Malito, Bhuta, Umbach, *supra* n 157, 503–512; Restrepo Amariles and McLachlan, *supra* n 91; Broome, Homolar, Kranke, *supra* n 48; W. Espeland, 'Narrating Numbers', in R. Rottenburg, S. Engle Merry, S-J. Park, J. Mugler (eds), *The Making of Governmental Knowledge through Quantification* (Cambridge: CUP, 2015) 56–75; Broome and Quirk, *supra* n 6; Merry, *supra* n 45, 27–43, 212–216; J. Snyder and A. Cooley, 'Conclusion: Rating the ratings craze: From consumer choice to public policy outcomes', in Cooley and Snyder (eds), *supra* n 17, 178–193; Jerven, *supra* n 49; Davis, Kingsbury, Merry, *supra* n 111; Krever, *supra* n 94; Kerry Rittich, 'Governing by Measuring. The Millennium Development Goals in Global Governance', in H. Ruiz Fabri, R. Wolfrum, J. Gogolin (eds), *Selected Proceedings of the European Society of International Law* (Oxford-Portland: Hart, 2010) 463–487; Espeland and Sauder, *supra* n 25.

mechanical recordings) that are defined by the metrics themselves as the only relevant evidence. With respect to evidence too, whenever a measure incorporates some technological elements, the overall rigidity of the system increases. If performance has to be proved by filling out a form, which is then automatically checked by a machine, any failure to fill out the form or by the machine reading equates to a bad performance, no matter how good performance actually was. Additional side-effects of this emphasis on form and evidence are a rise in the costs of producing, checking and storing evidentiary requirements, and the ease with which the system can be gamed or manipulated by anyone who is sufficiently smart or well-resourced to do so.

All of the methodological elements mentioned so far—the choice about which variables should be counted, and how, relying on which evidence, provided by whom, and so on and so forth—are important in determining the overall effects of measurements.[174] It has been noted above that these choices are influenced by empirical and practical constraints as much as they are influenced by assumptions about what good/bad performance means, what factors are causally relevant in determining improvement, and what ideal world the measurement strives to realize.[175] From this point of view, any performance-based measurement embodies a specific stance vis-à-vis the measured phenomena that is determined by both the technical and political constraints that affect those who conceive of and manage the measures. The less transparent a measure is as regards its underlying choices, the more difficult it will be to unpack and contest its results, all the more so where the measurement is expressed in statistical language or performed automatically by algorithms and codes (whether copyrighted or not).[176]

The technicality characterizing analog and automated performance-based measurements has other consequences. As specialized and complex artifacts, measures typically require, and imply the formation of, a technocracy capable of conceiving them and a bureaucracy capable of managing them. Experts in social quantification are usually needed in order to conceive the quantitative assessment; after it has been established, measures have to be repeated, the methodology has to be fine-tuned, and the results have to be interpreted and packaged in a publishable format by dedicated personnel. Specialized personnel may also be required to manage the procedures (whenever present) for providing interested parties with rights of access,

173 See section 5.2. How Metrics Calculations are Performed.

174 See section 5.3. Whose Metrics?.

175 See section 5.3. Whose Metrics?.

176 Among many, see Burk, *supra* n 62, 301–302; Catá Backer, *supra* n 38, 166; Selbst and Barocas, n 171, 1093–1094; Pasquale, *supra* n 39, 8, 16; Grimmelmann, *supra* n 165, 1736.

participation and review. However, the creation and involvement of a technocratic bureaucracy entail specific effects. On the one hand, the fact that many fundamental methodological choices are left in the hands of statisticians, programmers and data scientists de-politicizes decisions that may deserve to be politically negotiated, and adds a further layer of opacity to the measuring process, because technical decisions often lie beyond the comprehension of laymen. On the other hand, the need for bureaucracy in order to manage metrics empowers invisible and politically unaccountable personnel, and establishes a managerial mode of governance in which issues of substance and methodology are transformed into routinized controls to check that procedures and forms are respected. The side-effects produced by this managerial rule by experts may be limited when the empowerment of bureaucrats occurs within well-entrenched institutions whose conduct is constrained by clear administrative rules and practices. However, when bureaucratization occurs in the context of measures set up by other actors—such as international organizations, NGOs, and corporations—the power of technocracy may be constrained by nothing and may be accountable to none.[177]

To sum up, measuring performance as a regulatory technique is grounded on the self-explanatory power of allegedly objective evidence. Through compliance-based analysis, performance quantification empowers invisible technocracies, while also shaping, conveying and enforcing standards of behavior that are hardly visible, hardly negotiable, and hardly challengeable.

7 Comparative Law Matters

None of the above suggests that performance quantification produces the same effects always and everywhere.

177 This point too has been noted by many: see, among others, de Souza, *supra* n 61, 112–162; Catá Backer and McQuilla, *supra* n 37, 16; Catá Backer, *supra* n 38, 166–167; D. McGrogan, 'Human Rights Indicators and the Sovereignty of Technique', *European Journal of International Law* 27 (2016) 385–408; Merry, *supra* n 45, 27–35; Snyder and Cooley, *supra* n 172, 178; Davis, Kingsbury, Merry, *supra* n 111, 19–21; Cassese and Casini, *supra* n 60, 465–466; Restrepo Amariles, *supra* n 60, 12–14; J. Shkabatur, 'A Global Panopticon? The Changing Role of International Organizations in the Information Age', *Michigan Journal of International Law* 33 (2014) 159–214 (2011).

We have already highlighted that similar measures affect the social world differently, including the daily management of the law, depending on who carries them out, how, for the benefit of whom and for what purposes.[178] As noted at the beginning of this article,[179] measuring justice is different from assessing research quality; evaluating consumer creditworthiness is one thing, profiling consumption preferences is another; none of these has much in common with measures of states' and companies' efforts to achieve sustainability. However, the variety we are referring to goes deeper than sectoral heterogeneity.

The point is that each legal culture, system and order internalizes and reacts to the rise of performance-based quantification in its own way. Consider for instance of the varied response to the DB reports. All divisions of the WB readily embraced the DB, while many other international organizations attacked it;[180] several countries took the DB seriously, adopting legal reforms to comply with its requirements; others attempted to game it or simply cheated; still others strenuously tried to resist it.[181] Other examples may readily follow. For instance, measures of justice and research encroach upon judges' and researchers' independence to different extents in continental Europe, in common law jurisdictions and in East Asia.[182] In the wake of the Covid-19 pandemic, many East Asian countries (such as China, South Korea, Singapore, India), as well as Israel, have resorted to a variety of track and trace tools to monitor and nudge people's behavior, which were regarded with suspicion in the West.[183] The embrace of social credit in China has triggered fears of governmental surveillance in the West, where the same activities of gathering and relying on quantified performance data are often managed by corporate actors.[184] The Chinese decision to geo-block access from users in mainland China to a number of Western websites has successfully blocked many of the quantification activities routinely carried out by tech companies in the West, although the

178 See sections 5.1. Where Metrics Can Be Found–5.5. What Effects Metrics Trigger.

179 See section 1. Introduction.

180 See the illustrations provided *supra* n 78, 81–82.

181 See *supra* n 77, 79–80, 83, as well as the case study provided in M. Infantino, 'Measuring (the Effects of) Measurements: Four Global Legal Indicators in Italy', *Italian Law Journal* 6 (2020) 431–461.

182 See the references *supra* n 25.

183 Cf Liu, *supra* n 25; Angiolini, *supra* n 25; Bradford, Aboy, Liddell, *supra* n 25.

184 Infantino and Wang, *supra* n 38, 51–53; Devereaux and Peng, *supra* n 89, 7; Mac Síthigh and Siems, *supra* n 38, 1036–1046.

very same activities are in practice performed by local actors.[185] Even within the same legal tradition, corporate measures of people's creditworthiness, reliability, preferences and habits are permitted to different extents. To take an example from performance-based measures in the insurance sector in the West, the use of behavioral-based data in health and life insurance—whereby data concerning customers' behavior and performance are collected through connected devices and used to 'personalize' prices and contractual terms, adjusting them in real time—is for the time being remarkably different in the Anglo-American world than it is in continental Europe. In common law jurisdictions, many have noted an increasing reliance by insurance companies on automated personalization and have objected to the risks that this creates for individuals, allowing insurers to offer products and services at varying prices and to drive up insurance rates (or, where legal, to deny coverage) for riskier customers.[186] If one considers the continental European market, one finds that personalized insurance products are emerging much more slowly.[187] It is quite clear that that this divergence is due to differences not only in technological advancements and insurance market structures, but also in legal features such as the size of the welfare state and the strength of regulatory oversight over consumer protection and data sharing.[188]

All of the above shows that local legal features matter in determining how performance quantification is used as a regulatory technique. Although, as far

185 Bradford, *supra* n 73, 78–79, 203–204; for an updated list of the websites that cannot be accessed from mainland China, see ltl-beijing.com/websites-banned-in-china/ (retrieved 7 December 2023). With regard to AI-powered measurements, the proposal for a regulation of artificial intelligence currently under discussion in the European Union includes a few prohibitions of using AI for scoring and predicting human performances: see for instance points (c) and (d) of Article 5(1) of the 'Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)', 2021/0106(COD), June 2023 draft, artificialintelligenceact.eu/wp-content/uploads/2023/08/AI-Mandates-20-June-2023.pdf (retrieved 7 December 2023).

186 B. Soyer, 'Use of Big Data Analytics and Sensor Technology in Consumer Insurance Context: Legal and Practical Challenges', *Cambridge Law Journal* 81 (2022) 165–194, esp. at 186; Zuboff, *supra* n 28, 213; McGurk, *supra* n 63; Lupton, *supra* n 28, 123; O'Neil, *supra* n 1, 161–178.

187 M. Infantino, 'Automated Personalization and Consumer Insurtech in European Law: Prospects and Challenges', in P. Tereszkiewicz and C. Poncibò (eds), *European Insurance Contract Law in the Age of Digitalization* (Cham: Springer, forthcoming 2024); Jeanningros and McFall, *supra* n 120, 4–5; G. Meyers and I. Van Hoyweghen, "Happy failures': Experimentation with behavior-based personalisation in car insurance', *Big Data and Society* (2020) 1–14.

188 See the authors cited *supra* n 187.

as we are aware, no legal system has adopted a set of rules specifically designed for performance-based metrics, a number of substantial and procedural obligations may arise from a variety of existing bodies of law, including constitutional and administrative law, anti-discrimination provisions, privacy regulation and consumer protection laws. Moreover, many legal systems have enacted bans or imposed specific limitations on activities that are directly associated with performance-based measurements. A simple example comes from the rules that in many places forbid the collection of census data concerning, e.g., sex, religion and ethnicity.[189] As far as selective limitations are concerned, several countries have adopted rules that aim to limit conflicts of interest and to foster transparency in the credit rating industry.[190] In the US, the Fair Credit Reporting Act requires any person or body that uses information collected by consumer reporting agencies for credit, insurance, or employment purposes to notify the affected person when any adverse action is taken on the basis of those reports.[191] In the European Union, online platforms must now disclose to users, in an easily comprehensible manner, the parameters and criteria (including those related to the tracking and profiling of users' behavior) according to which the information contained in the platform is prioritized and presented to them.[192] These and other similar rules clearly influence what is measured and how, by whom and for what purposes.

189 Merry, *supra* n 45, 14; L.-G. Tin, 'Qui a peur des statistiques ethniques', in I. Bruno, E. Didier, J. Prévieux (eds), *Stat-Activisme. Comment lutter avec des nombres* (Paris: La Découverte, 2014) 155–166; W. Seltzer and M. Anderson, 'The Dark Side of Numbers. The Role of Population Data Systems in Human Rights Abuses', *Social Research* 68 (2001) 481–507, at 495.

190 Cp. the Securities Exchange Act of 1934, Section 15E(d) (15 U.S.C.A. 7807) (as amended) in the US, Regulation (EC) No 1060/2009 of the European Parliament and of the Council of 16 September 2009 on credit rating agencies (as amended) in the EU, and the Interim Measures for the Administration of the Credit Rating Industry adopted in 2019 in China. On the many limitations of these regulations, see Bussani, *supra* n 142.

191 Fair Credit Reporting Act, 15 U.S.C. §§ 1681–1681x.

192 See Article 27 of the Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act). According to Article 93, the Regulation will apply from February 17, 2024. Again in the European Union, the above-mentioned proposal for a regulation of artificial intelligence imposes several obligations (concerning risk management, data governance, transparency, human oversight and accuracy: Articles 8–15 of the 'Proposal', *supra* n 185) on producers and users of (what the proposal defines as) high-risk AI systems, many of which (are classified as such exactly because they) include mechanisms for measuring performance: see for instance Annex III, nos 4(b), 5(a), 5(b), 6(a), 'Proposal', *supra* n 185).

As a reminder of the fact that Western-based metrics often aim to measure-and-control the world,[193] one final note is in order. Assessing (if not plainly promoting) Western-style rule of metrics would require a fully-fledged understanding of the variety of historical, economic, and cultural backgrounds against which the different legal systems flourish, converge and diverge, compete with, and imitate one another. Western initiatives should be brought forward with a view that should be inclusive of Western as well as of non-Western rationales—those coming from their respective social, political, economic, legal traditions and frameworks. In other words, it would be necessary to search for criteria that can calibrate judgments and options concerning the variable standards that other legal experiences offer, rather than concerning the measure of a self-established messianic spirit. Such a spirit is all the more unhelpful and even dysfunctionally absurd when coupled with an unquestioning reliance on the virtues of metrics. This is why any serious analysis of whether and how to spread the rule of metrics should tap into reservoirs of knowledge different from these utilitarian and simplified toolkits. Theory and practice need to take on board the lessons of anthropology, history, and comparative law. These are not mere academic disciplines, or a sort of intellectual exercise devoid of any actual impact on the adoption and processing of legal rules. On the contrary, these fields of study are powerful heuristic tools capable of driving the analysis towards solutions that promise to work as they are (in principle, at least[194]) unfettered by Western-centered biases and because they acknowledge what matters in a local setting.[195]

It follows from all the above that comparative lawyers may have an important role to play in placing their knowledge about legal cultures, legal transplants, and legal change at the service of the establishment and in offering a critique of rule of metrics. Furthermore, studying from a comparative perspective how legal systems, orders and cultures are variedly embracing the quantitative turn would (not only fill a gap in legal research vis-à-vis other social sciences, but

193 See section 3. Instances of Legal Metrics.

194 On the Western biases that affect some purported comparative law scholarship, see M. Bussani and U. Mattei, 'Diapositives Versus Movies: The Inner Dynamics of the Law and Its Comparative Account', in M. Bussani and U. Mattei (eds), *Cambridge Companion to Comparative Law* (Cambridge: CUP, 2012) 3, 4–7; G. Frankenberg, 'The Innocence of Method—Unveiled: Comparison as an Ethical and Political Act', *Journal of Comparative Law* 9 (2014) 222.

195 See also J.H. Merryman, 'Comparative Law and Social Change: On the Origins, Style, Decline and Revival of the Law and Development Movement', *American Journal of Comparative Law* 25 (1977) 457; L.M. Friedman, 'Legal Culture and Social Development', *Law and Society Review* 4 (1969) 29.

also) help understand a governance and regulatory dynamic that is largely invisible, largely unaccountable, and increasingly powerful.

8 Conclusions

This paper has shown how ubiquitous quantification is. We live in societies where there is widespread collection of information, which is then used by different actors at various levels to make decisions that materially affect our everyday life. The shift towards quantification as a regulatory technique is undoubtedly here to stay.

What our research has also shown is that legal metrics are on the rise everywhere, although the design and effects are uneven. Measures are made, used and consumed in different ways, and measures can similarly be controlled in different ways. The burgeoning variety of performance-based assessments calls for more comparative research into legal metrics. We need to understand more about which forms of quantitative measures are widespread, in which sectors and regions, made by whom and producing what regulatory effects. We also need to understand more about the existing and emerging legal rules that may apply to these measures, and the extent to which they affect the contents, style and impact of social quantification. This research is both urgent and topical.

We are aware that many of the arguments proposed in this paper may not be acceptable to all readers. We know that greater precision is required in the definition of performance, processes/people/organizations, legal effects, and regulation. We know that much more empirical information is needed about the spread of quantification of the social world, as well as more fine-grained attempts to distinguish between different kinds of quantitative initiatives. We know that readers may have understandings of legal sources that differ from the one adopted here, and may resist the idea of equating measurements to regulation—and even more so to law. However, irrespective of the many refinements that may be proposed to all of the above, and no matter whether one (dis)agrees with the labels and qualifications used throughout the paper, we hope to have demonstrated that performance-based measurements often produce regulatory effects that affect, and are affected by, the legal contexts in which they operate.