Governing the world at a distance: the practice of global benchmarking

ANDRÉ BROOME AND JOEL QUIRK*

Abstract. Benchmarking practices have rapidly diffused throughout the globe in recent years. This can be traced to their popularity amongst non-state actors, such as civil society organisations and corporate actors, as well as states and international organisations (IOs). Benchmarks serve to both 'neutralise' and 'universalise' a range of overlapping normative values and agendas, including freedom of speech, democracy, human development, environmental protection, poverty alleviation, 'modern' statehood, and 'free' markets. The proliferation of global benchmarks in these key areas amounts to a comprehensive normative vision regarding what various types of transnational actors should look like, what they should value, and how they should behave. While individual benchmarks routinely differ in terms of scope and application, they all share a common foundation, with normative values and agendas being translated into numerical representations through simplification and extrapolation, commensuration, reification, and symbolic judgements. We argue that the power of benchmarks chiefly stems from their capacity to create the appearance of authoritative expertise on the basis of forms of quantification and numerical representation. This politics of numbers paves the way for the exercise of various forms of indirect power, or 'governance at a distance', for the purposes of either status quo legitimation or political reform.

André Broome is Associate Professor of International Political Economy in the Department of Politics and International Studies and Director of the Centre for the Study of Globalisation and Regionalisation, University of Warwick. His publications include *The Currency of Power* (Palgrave, 2010) and *Issues and Actors in the Global Political Economy* (Palgrave, 2014).

Joel Quirk is Associate Professor in the Department of Political Studies, University of the Witwatersrand. His publications include *The Anti-Slavery Project* (Penn, 2011) and *Mobility Makes States* (Penn, 2015). Joel is a current member of the International Scientific Committee of the UNESCO Slave Route Project, and is also an *openDemocracy* editor for Beyond Trafficking and Slavery.

Introduction

Global benchmarking comprises a distinct type of transnational practice in contemporary world politics, which involves the development and application of

comparative metrics of performance. While benchmarking is not in itself a new phenomenon, the last three decades have been marked by a sharp increase in the density, complexity, and coverage of global benchmarking practices.[1] Much of this ongoing trend can be traced to the globalisation of an 'audit explosion' that began in the 1980s in domestic political contexts, and which has had far-reaching ramifications for both public and private processes of transnational governance.[2] Other key contributing factors include the rapid proliferation of non-governmental organisations (NGOs) in areas such as human rights, health, gender, and the environment, together with a parallel shift from state to private regulation at a corporate level.[3] Even the 'ivory tower' of academia is increasingly governed through ratings, rankings, and measurements of how well higher education institutions perform in comparison to their competitors.[4] These and other developments have not only dramatically expanded the pool of prospective 'benchmarkers'. They have also fostered an environment where benchmarks have gained considerable legitimacy and authority.

In its most basic form, benchmarking involves the classification of relative performance or value. In this article and for the Special Issue, benchmarking is used as an umbrella term for a wide range of comparative evaluation techniques – such as audits, rankings, indicators, indexes, baselines, or targets – which systematically assess the performance of actors, populations, or institutions on the basis of standardised measurements, metrics, and rankings. More specifically, benchmarking involves one or more of the following forms of comparative assessment: (1) quality of *conduct*, or how well actors have discharged their responsibilities in specific areas; (2) quality of *design*, or how well specific policies, laws, or institutions have been formulated and applied; and (3) quality of *outcomes*, or how well activities in specific areas align with defined goals (irrespective of who is actually responsible for the overall outcomes).

In this article we identify and analyse a number of core features of benchmarking as a distinct mode of governance in world politics. We begin our analysis by locating global benchmarking within an emerging literature that focuses on how and why both states and non-state actors have sought to regulate and shape transnational issues through indirect forms of power, rather than through direct compulsion. Building upon this literature, we argue that benchmarking can be best understood as an exercise in 'governing at a distance', wherein the power of benchmarks primarily stems from their capacity to indirectly shape procedural standards, issue expertise, institutional

---

[1] See the Global Benchmarking Database (N = 205), version 1.8, available at: {www.warwick.ac.uk/globalbenchmarking/database} accessed 5 June 2015. We are grateful to Matthias Kranke for research assistance with compiling data on Global Benchmarks.

[2] See, for example, Michael Power, *The Audit Society: Rituals of Verification* (Oxford: Oxford University Press, 1997); Wendy Nelson Espeland and Mitchell L. Stevens, 'A sociology of quantification', *European Journal of Sociology*, 49:3 (2008), pp. 401–36; Michael Power, 'Evaluating the audit explosion', *Law and Policy*, 25:3 (2003), pp. 185–202.

[3] See, for example, Richard Locke, *The Promise and Limits of Private Power: Promoting Labor Standards in a Global Economy* (Cambridge: Cambridge University Press, 2013); Anne-Marie Slaughter, *A New World Order* (Princeton: Princeton University Press, 2005); Samuel Moyn, *The Last Utopia: Human Rights in History* (Cambridge, MA: Harvard University Press, 2012); Tim Büthe and Walter Mattli, *The New Global Rulers: The Privatization of Regulation in the World Economy* (Princeton: Princeton University Press, 2011); Angelina Fisher, 'From diagnosing under-immunization to evaluating health care systems: Immunization coverage indicators as a technology of global governance', in Kevin E. Davis, Angelina Fisher, Benedict Kingsbury, and Sally Engle Merry (eds), *Governance by Indicators: Global Power through Quantification* (Oxford: Oxford University Press, 2012), pp. 217–46.

[4] Jack Snyder and Alexander Cooley, 'Conclusion: Rating the ratings craze: From consumer choice to public policy outcomes', in Alexander Cooley and Jack Snyder (eds), *Ranking the World: Grading States as a Tool of Global Governance* (Cambridge: Cambridge University Press, 2015), pp. 180–2.

obligations, and political conversations. Much of the power of benchmarking is bound up in the mechanics and effects of ranking and quantification, which in turn generate a form of 'constructed objectivity' that acts back upon the reality it aims to describe.[5] The recent popularity of benchmarks can also be traced to their capacity to promote otherwise highly contentious policy goals and political agendas by means of rhetorical appeals to the ostensibly neutral language of technocratic assessment and numerical comparison. Complex social phenomena become legible by means of quantification, extrapolation, and simplification. Concepts such as freedom, development, and democracy, which academics routinely describe as essentially contested, instead appear as fixed, unproblematic, and reified categories.

We have divided this article into four main sections. The first section briefly situates our approach to global benchmarking within the larger context of existing literatures in International Relations (IR) on political activism and norms, rational design and institutions, and governmentality and expertise. In the second section, we focus upon the mechanics and effects associated with translating normative values into numerical representations. By radically reducing issue complexity, benchmarks have the potential to alter 'how people think about things and how information moves around the world'.[6] This process of translation can be divided into a series of steps common to all forms of benchmarking: simplification and extrapolation, commensuration, reification, and symbolic judgment. The third section examines the political ramifications of these processes of quantification and numerical representation for transnational governance, along with the political impact of the alignment between benchmarks and other agendas. The final section, which introduces a typology of global benchmarking practices, develops this line of inquiry further. We divide global benchmarking practices into four main categories: (1) statecraft; (2) international governance; (3) private market governance; and (4) transnational advocacy. We conclude the article by identifying a series of core questions for a new research agenda on global benchmarking in International Relations.

Governing at a distance: Benchmarking and IR theory

We understand global benchmarking as a mode of *transnational governance*, which comprises a patchwork of political structures within and above the state that envelope, constrain, and enable various actors. Drawing on Marie-Laure Djelic and Kerstin Sahlin-Andersson's definition, the boundaries of the 'transnational' arena stretch beyond the jurisdiction of domestic governance structures and are not limited to one specific region.[7] Benchmarking practices are *global* when they aim to produce comparative measurements of performance across numerous countries and regions. The units of analysis comprise a range of transnational actors, such as states,

---

[5] Peter Berger and Thomas Luckmann, *The Social Construction of Reality: A Treatise in the Sociology of Knowledge* (London: Allen Lane, 1967), p. 78.

[6] Wendy Nelson Espeland and Michael Sauder, 'The dynamism of indicators', in Kevin E. Davis, Angelina Fisher, Benedict Kingsbury, and Sally Engle Merry (eds), *Governance by Indicators: Global Power through Quantification* (Oxford: Oxford University Press, 2012), p. 91.

[7] Marie-Laure Djelic and Kerstin Sahlin-Andersson, 'Introduction: a world of governance – the rise of transnational regulation', in Marie-Laure Djelic and Kerstin Sahlin-Andersson (eds), *Transnational Governance: Institutional Dynamics of Regulation* (Cambridge: Cambridge University Press, 2008), p. 4; see also Emanuel Adler and Vincent Pouliot, 'International practices: Introduction and framework', in Emanuel Adler and Vincent Pouliot (eds), *International Practices* (Cambridge: Cambridge University Press, 2011), pp. 7–8.

international organisations, or corporate subsidiaries within a global production network. A global framework applies even if some countries or actors are excluded.

Many global benchmarking efforts have focused on the economic and political performance of states. Early examples on the economic front include gross domestic product and the System of National Accounts, developed during the 1930s in the United States.[8] With respect to political performance, pioneering examples of benchmarking include the standardised international monitoring of elections and the annual 'Freedom in the World' rankings published by Freedom House (an NGO part-funded by the US government) since 1973.[9] In addition to measuring cross-national economic and political performance, benchmarking has also become an important means for evaluating corporate performance. This involves systematic comparisons to evaluate individual firm competitiveness and to establish industry 'best practice' processes based on measures of quality, time, and cost.[10] This form of benchmarking extends to commercially motivated efforts to evaluate market conditions, financial performance, and creditworthiness, most notably by means of credit ratings. In some cases political and commercial concerns have been integrated, such as in the political and country risk ratings published by the PRS Group since 1980.[11] One of the distinctive features of corporate benchmarking is that it frequently takes the form of self-benchmarking against peers with a view to improving, validating, or refining overall performance and internal processes,[12] which is broadly comparable to the use of benchmarking by individual states for the purposes of domestic governance.

Our main focus here is on benchmarking by external transnational actors, rather than internal self-benchmarking. Some notable examples of this trend include measures of state performance in relation to international human rights obligations,[13] global indexes of country 'competitiveness',[14] measurements of the perception of corruption in state institutions,[15] assessments of democratic freedom and the

---

[8] Yoshiko M. Herrera, *Mirrors of the Economy: National Accounts and International Norms in Russia and Beyond* (Ithaca, NY: Cornell University Press, 2010); Lorenzo Fioramonti, *Gross Domestic Problem: The Politics Behind the World's Most Powerful Number* (London: Zed Books, 2013); Daniel Mügge, 'Fickle formulas: Towards a political economy of macroeconomic measurements', *Journal of European Public Policy*, forthcoming.

[9] Judith G. Kelley, *Monitoring Democracy: When International Election Observation Works, and Why It Often Fails* (Princeton: Princeton University Press, 2012); see Alexandra Homolar, 'Human security benchmarks: Governing human wellbeing at a distance', *Review of International Studies*, 41:5 (2015), pp. 843–63.

[10] Wendy Larner and Richard Le Heron, 'Global benchmarking: Participating "at a distance" in the globalizing economy', in Wendy Larner and William Walters (eds), *Global Governmentality: Governing International Spaces* (Abingdon: Routledge, 2004), pp. 212–32.

[11] Nikhil K. Dutta, 'Accountability in the generation of governance indicators', in Kevin E. Davis, Angelina Fisher, Benedict Kingsbury, and Sally Engle Merry (eds), *Governance by Indicators: Global Power through Quantification* (Oxford: Oxford University Press, 2012), pp. 437–64.

[12] See Genevieve LeBaron and Jane Lister, 'Benchmarking global supply chains: the power of the "ethical audit" regime', *Review of International Studies*, 41:5 (2015), pp. 905–24.

[13] Kate Raworth, 'Measuring human rights', *Ethics and International Affairs*, 15:1 (2001), pp. 111–31; Sakiko Fukuda-Parr, 'Millennium development goal 8: Indicators for international human rights obligations?', *Human Rights Quarterly*, 28:4 (2006), pp. 966–97.

[14] Tore Fougner, 'Neoliberal governance of states: the role of competitiveness indexing and country benchmarking', *Millennium: Journal of International Studies*, 37:2 (2008), pp. 303–26.

[15] Peter Larmour, 'Civilizing techniques: Transparency international and the spread of anti-corruption', in Brett Bowden and Leonard Seabrooke (eds), *Global Standards of Market Civilization* (Abingdon: Routledge, 2006), pp. 95–106; Laura Langbein and Stephen Knack, 'The worldwide governance indicators: Six, one, or none?', *Journal of Development Studies*, 46:2 (2010), pp. 350–70; Paul M. Heywood and Jonathan Rose, '"Close but no cigar": the measurements of corruption', *Journal of Public Policy*, 34:3 (2014), pp. 507–29.

transparency of elections,[16] headcount measures of absolute poverty,[17] and measures of state 'fragility'.[18] Such external benchmarking by transnational actors has rapidly proliferated around the world over the last three decades.

IR theorists have developed a number of insights and arguments that can be usefully applied in order to better understand the politics of benchmarking. Since relatively few IR theorists have focused upon benchmarking as a specific object of analysis,[19] we briefly engage with a number of allied literatures that speak to similar and related topics, most notably in relation to theories of norms and human rights, rational design and cooperation, and governmentality. Over the last two decades, IR theorists have repeatedly demonstrated that normative arguments and collective identities have generated outcomes that cannot be explained in terms of power and interest alone.[20] This has in turn resulted in sustained interest in the techniques, alliances, and arguments employed by 'agents of change'. Many of the political levers that theorists have identified – such as reputational challenge, communicative networks, and patterns of socialisation – can also be applied to the politics of benchmarking, particularly in relation to transnational advocacy. Especially relevant is the emerging literature on 'merchants of morality', which seeks to explain why and how some issues have become subject to mobilisation while others remain dormant;[21] why some political causes and organisations have secured greater success (or 'salience') than their competitors;[22] and how the accumulation and application of 'credibility' has emerged as a key source of authority and influence for NGOs.[23] Within the context of this recent literature, benchmarking can be at least partially theorised in terms of the larger dynamics of market competition between political causes and organisations for resources, audiences, allies, and credibility.

Much of the recent proliferation of global benchmarks can be traced to their perceived capacity to help build the reputation of specific organisations as

---

[16] Diego Giannone, 'Political and ideological aspects in the measurement of democracy: the Freedom House case', *Democratization*, 17:1 (2010), pp. 68–97.

[17] Antje Vetterlein, 'Seeing like the World Bank on poverty', *New Political Economy*, 17:1 (2012), pp. 35–58.

[18] Nehal Bhuta, 'Governmentalizing sovereignty: Indexes of state fragility and the calculability of political order', in Kevin E. Davis, Angelina Fisher, Benedict Kingsbury, and Sally Engle Merry (eds), *Governance by Indicators: Global Power Through Quantification* (Oxford: Oxford University Press, 2012), pp. 132–62.

[19] See the various contributions to Alexander Cooley and Jack Snyder (eds), *Ranking the World: Grading States as a Tool of Global Governance* (Cambridge: Cambridge University Press, 2015); see also Judith G. Kelley and Beth A. Simmons, 'Politics by number: Indicators as social pressure in International Relations', *American Journal of Political Science*, 59:1 (2015), pp. 55–70.

[20] See, for example, Thomas Risse, Stephen Ropp, and Kathryn Sikkink (eds), *The Power of Human Rights: International Norms and Domestic Change* (Cambridge: Cambridge University Press, 1999); Martha Finnemore and Kathryn Sikkink, 'International norm dynamics and political change', *International Organization*, 52:4 (1998), pp. 887–917; Rawi Abdelal, Mark Blyth, and Craig Parsons (eds), *Constructing the International Economy* (Ithaca, NY: Cornell University Press, 2010); Susan Park and Antje Vetterlein (eds), *Owning Development: Creating Policy Norms in the IMF and the World Bank* (Cambridge: Cambridge University Press); Alexander Betts and Phil Orchard (eds), *Implementation and World Politics: How International Norms Change Practices* (Oxford: Oxford University Press, 2014).

[21] See, for example, Charli Carpenter, 'Studying issue (non)-adoption in transnational advocacy networks', *International Organization*, 61:3 (2007), pp. 643–67; Charli Carpenter, 'Setting the advocacy agenda: Issues and non-issues around children and armed conflict', *International Studies Quarterly*, 51:1 (2007), pp. 99–120.

[22] See Charli Carpenter, *Lost Causes: Agenda-Setting and Agenda-Vetting in Global Issue Networks* (Ithaca, NY: Cornell University Press, 2014); Clifford Bob, *The Marketing of Rebellion: Insurgents, Media, and International Activism* (Cambridge: Cambridge University Press, 2005); Wendy Wong, *Internal Affairs: How the Structure of NGOs Transforms Human Rights* (Ithaca: Cornell University Press, 2012).

[23] Peter A. Gourevitch, David A. Lake, and Janice Gross Stein (eds), *The Credibility of Transnational NGOs: When Virtue is Not Enough* (Cambridge: Cambridge University Press, 2012); L. David Brown, *Creating Credibility: Legitimacy and Accountability for Transnational Civil Society* (London: Kumarian Press, 2008).

'issue experts'.[24] The popularity of benchmarking as a strategic tool for producing authoritative expertise – or at least the public appearance of expertise – is most notable in relation to NGOs and some IOs, which frequently find themselves in competition with their peers for allies, attention, and resources.[25] Thanks to the digital revolution of the last two decades,[26] it is often cheaper and easier to formulate and disseminate benchmarks than to engage in most forms of on the ground intervention. These conditions have contributed to an increasing level of market saturation, with NGOs, IOs, and other actors launching competing benchmarks as part of strategic efforts to create and consolidate a distinctive brand.

It is also important to take into account the intersections between expertise, authority, and indirect power. Over the last decade, a number of IR scholars have focused on the role of expert knowledge in the exercise of indirect power.[27] Recent works have demonstrated that expert knowledge and authority have helped to shape the architecture and practice of transnational governance,[28] to construct authority at the transnational level and effect transnational decision-making,[29] and to configure incentive systems that drive the global diffusion of common policy models and normative standards.[30]

The literature on rational design and institutional choice also offers further insights into benchmarking practices, most notably in relation to cooperation, coordination, and regulation. Rationalist theories can be loosely grouped together around the basic idea that both state and non-state actors often have a mutual interest in coordinating and codifying their activities across different spheres of global governance, and that these interests help to explain variations in the design and operation of international institutions. These overlapping interests may include shared efficiency gains, similar interests associated with information sharing and standardisation, and the mutual benefits gained from institutional arrangements that overcome collective action problems. This final point is based on the

---

[24] Peter J. May, Chris Koski, and Nicholas Stramp, 'Issue expertise in policymaking', *Journal of Public Policy*, DOI: 10.1017/S0143814X14000233; see also Leonard Seabrooke, 'Epistemic arbitrage: Transnational professional knowledge in action', *Journal of Professions and Organizations*, 1:1 (2014), pp. 49–64.

[25] Ellen Gutterman, 'The legitimacy of transnational NGOs: Lessons from the experience of transparency international in Germany and France', *Review of International Studies*, 40:2 (2014), pp. 391–418; see also Leonard Seabrooke and Duncan Wigan, 'How activists use benchmarks: Reformist and revolutionary benchmarks for global economic justice', *Review of International Studies*, 41:5 (2015), pp. 887–904. Ole Jacob Sending, *The Politics of Expertise: Competing for Authority in Global Governance* (Ann Arbor: University of Michigan Press, 2015), p. 12.

[26] Lance Bennett and Alexandra Segerberg, *The Logic of Connective Action: Digital Media and the Personalization of Contentious Politics* (Cambridge: Cambridge University Press, 2014).

[27] This has also been a major theme of work in cognate fields, such as Law. See Kevin E. Davis, Benedict Kingsburgy, and Sally Engle Merry, 'Introduction: the local-global life of indicators: Law, power, and resistance', in Sally Engle Merry, Kevin E. Davis, and Benedict Kingsbury (eds), *The Quiet Power of Indicators: Measuring Governance, Corruption, and Rule of Law* (Cambridge: Cambridge University Press, 2015), pp. 1–24.

[28] Oded Löwenheim, 'Examining the state: a Foucauldian perspective on international "governance indicators"', *Third World Quarterly*, 29:2 (2008), pp. 255–74; Jacqueline Best, *Governing Failure: Provisional Expertise and the Transformation of Global Development Finance* (Cambridge: Cambridge University Press, 2014); Michael Barnett and Martha Finnemore, *Rules for the World: International Organizations in Global Politics* (Ithaca, NY: Cornell University Press, 2004).

[29] Cornel Ban, *Ruling Ideas: How Global Economic Paradigms Go Local* (Oxford: Oxford University Press, 2016); André Broome and Leonard Seabrooke, 'Shaping policy curves: Cognitive authority in transnational capacity building', *Public Administration*, doi: 10.1111/padm.12179; Sending, *The Politics of Expertise*.

[30] J. C. Sharman, 'Power and discourse in policy diffusion: Anti-money laundering in developing states', *International Studies Quarterly*, 52:3 (2008), pp. 635–56; Kelley and Simmons, 'Politics by number'.

understanding that 'states and other international actors, acting for self-interested reasons, design institutions purposefully to advance their joint interests'.[31]

These types of arguments help to explain why a wide range of transnational actors have increasingly embraced benchmarks and benchmarking. As we explore in more detail below, much of the appeal of benchmarks stems from their capacity to translate complex phenomena into numerical information. This makes it feasible for non-experts to make comparisons across a diverse range of cases and contexts, and enables the definition of targets and numerical criteria that can facilitate evaluations of relative performance. From this vantage point, the recent proliferation of benchmarking can be at least partially traced to a combination of rational interests, market demand, and institutional design. This is most notable in relation to private market governance, where one of the main motivations behind benchmarking has been to produce useable information that improves how actors respond to market forces and conditions.

In addition to information sharing, benchmarks play an increasingly central role when it comes to standardising and coordinating corporate policies on issues such as labour conditions and corporate social responsibility.[32] Similarly, benchmarking efforts now play a key role in policy coordination and institutional design among states and IOs faced with collective action problems over climate change, disaster management, and human development.[33] These efforts have not always been successful in bringing about desired outcomes (and there are circumstances when benchmarking can be used to deflect pressure for larger reforms), but it is nonetheless clear that there are a number of occasions when benchmarking can be theorised as a product of rational interests and cooperation.

While the existing IR scholarship on human rights and international institutions offers useful insights, these are not sufficient enough to fully understand the practices and politics of global benchmarking. We therefore draw inspiration from a growing literature concerned with governmentality, the exercise of indirect power, and related technologies of rule over distant entities in the international arena.[34] Governmentality has proved especially popular within IR circles as a way to theorise how forms of liberal or neoliberal governance have been able to exercise power at a distance by both constraining and channelling the social, political, and institutional horizons of specific actors and institutions.[35] One of the key points at issue here is 'how certain identities

---

[31] Barbara Koremenos, Charles Lipson, and Duncan Snidal, 'The rational design of international institutions', *International Organization*, 55:4 (2001), p. 781; see also Joseph Jupille, Walter Mattli, and Duncan Snidal, *Institutional Choice and Global Commerce* (Cambridge: Cambridge University Press, 2013).

[32] See LeBaron and Lister, 'Benchmarking global supply chains'; James Harrison and Sharifah Sekalala, 'Addressing the compliance gap? UN initiatives to benchmark the human rights performance of states and corporations', *Review of International Studies*, 41:5 (2015), pp. 925–45.

[33] See Tony Porter, 'Global benchmarking networks: the cases of disaster risk reduction and supply chains', *Review of International Studies*, 41:5 (2015); Liam Clegg, 'Benchmarking and blame games: Exploring the contestation of the Millennium Development Goals', *Review of International Studies*, 41:5 (2015); Caroline Kuzemko, 'Climate change benchmarking: Constructing a sustainable future?', *Review of International Studies*, 41:5 (2015), all in this Special Issue.

[34] Carl Death, 'Governmentality at the limits of the international: African politics and Foucauldian theory', *Review of International Studies*, 39:3 (2013), p. 768; see also Ole Jacob Sending and Iver B. Neumann, 'Governance to governmentality: Analyzing NGOs, states, and power', *International Studies Quarterly*, 50:3 (2006), pp. 651–72.

[35] Governmentality approaches are based on the work of Michel Foucault, see Fougner, 'Neoliberal governance of states'; Jonathan Joseph, 'The limits of governmentality: Social theory and the international', *European Journal of International Relations*, 16:2 (2010), pp. 223–46; Wanda Vrasti, 'Universal but not truly "global": Governmentality, economic liberalism, and the international', *Review of International Studies*, 39:1 (2013), pp. 49–69.

and action-orientations are defined as appropriate and normal and how relations of power are implicated in these processes'.[36] This theme has particular resonance in the case of global benchmarking, because benchmarks primarily operate by quantifying and projecting normative criteria regarding the parameters of appropriate conduct and performance.[37]

The literature on governmentality is especially useful for the insight that benchmarking functions to make diverse forms of behaviour legible and amenable to intervention.[38] However, existing applications of governmentality, which mostly analyse techniques of government in domestic contexts, cannot simply be stapled on to analyses of the transnational arena. Among other things, transnational governance initiatives are characterised by a high degree of variation in both the rate and the form of implementation across different jurisdictions.[39] Governmentality approaches are also less useful for understanding how benchmarking practices facilitate the political agendas of specific actors and organisations. We suggest that global benchmarks can nonetheless be usefully located within larger patterns of governmentality associated with contemporary transnational governance. These patterns are closely associated with indirect forms of power that establish appropriate standards of behaviour across a wide variety of policy domains.[40] Individual benchmarks tend to overlap in multiple ways, and therefore contribute to the diffusion of normative visions and agendas regarding what transnational actors should look like, what they should value, and how they should behave.

Translating normative values into numerical representations

Global benchmarking tends to be heavily reliant upon rhetorical appeals to authoritative expertise. Instead of relying upon forms of direct compulsion (actor *A* compels actor *B* to do what *A* wants),[41] global benchmarks usually operate by orienting how specific actors: (1) conceptualise their options, obligations, and opportunities; and (2) seek to legitimate and justify their performance and perceived relative standing. It is in within this context that benchmarking practices can be regarded as an exercise in governance at a distance, which combines indirect power, expert authority, and transnational governmentality. This also means that the political effects of benchmarking tend to be cumulative and subtle, rather than overt and immediate, but they can nonetheless have a major influence over processes of agenda-setting in transnational governance.

The recent proliferation of global benchmarks owes a major debt to the political and popular appeal of numbers as information shortcuts, whereby complex and

---

[36] Iver B. Neumann and Ole Jacob Sending, *Governing the Global Polity: Practice, Mentality, Rationality* (Ann Arbor: University of Michigan Press, 2010), p. 10; see also Mitchell Dean, *Governmentality: Power and Rule in Modern Society* (Thousand Oaks, CA: Sage, 2009).

[37] See also Katja Freistein, 'Effects of indicator use: a comparison of poverty measuring instruments at the World Bank', *Journal of Comparative Policy Analysis: Research and Practice*, DOI:10.1080/13876988.2015.1023053.

[38] See, for example, Larner and Le Heron, 'Global benchmarking'; Fougner, 'Neoliberal governance of states'.

[39] Joseph, 'The limits of governmentality', p. 243.

[40] Neumann and Sending, *Governing the Global Polity*, p. 66; Lorenzo Fioramonti, *How Numbers Rule the World: The Use and Abuse of Statistics in Global Politics* (London: Zed Books, 2014); Sending, *The Politics of Expertise*.

[41] Michael Barnett and Raymond Duvall, 'Power in international politics', *International Organization*, 59:1 (2005), pp. 39–75.

contested normative values are translated into simplified numerical representations. This process of translation not only helps to obscure their normative foundations, it also enables non-experts to make crude comparisons of relative performance regarding complex phenomena at a transnational level. This translation process is common to all forms of benchmarking, and can be divided into four distinct components:[42]

- simplification and extrapolation
- commensuration
- reification
- symbolic judgement

Simplification and extrapolation are preconditions of quantification. Simplification comes in many different forms, but the most common denominator is when complexity and contextual detail is 'lost in translation' in the pursuit of quantification and comparability. Since not every sphere of human activity can be easily quantified, benchmarking efforts have a tendency to gravitate towards behaviours that can be more easily and effectively translated into a numerical form, and thereby end up generating data that is chiefly based upon a narrow subset of contributing factors. Simplification also tends to overlook context-specific idiosyncrasies and histories in favour of an emphasis upon more general properties. The inherent limitations of simplification are often further complicated by extrapolation, which refers to efforts to 'plug the gaps' when available data falls short. Quantification requires reliable and comprehensive information, yet reliable information can often be in short supply in many contexts and countries. Faced with persistent and significant shortfalls, benchmarkers can end up extrapolating based upon what they already know – or what they think they know – which can result in highly speculative findings that later take on the imprimatur of facts once they are translated into numerical form.

Another component of numerical translation is commensuration, which refers to 'the expression or measurement of characteristics normally represented by different units according to a common metric'.[43] Otherwise dissimilar political, economic, and social conditions become easily comparable by translating qualities into quantities. Once qualities are translated into quantities, they can be graded and assessed in terms of their orders of magnitude. Commensuration therefore imposes a form of homogeneity among disparate entities that is imagined to be 'a property of the object rather than something produced by quantification'.[44] In addition, there are further advantages associated with the 'neutrality' and 'objectivity' commonly ascribed to numerical rankings and representations. 'Numbers are not like words, which require interpretation', but are instead widely perceived to present unbiased facts.[45] There is a widespread tendency to fixate on specific numerical claims, which create 'anchoring effects' by establishing referents that shape how people later conceptualise specific issues.[46] These 'anchoring effects' also underpin the capacity of

---

[42] Simplification and extrapolation, commensuration, reification, and symbolic judgement are distinct components of the larger process of global benchmarking, rather than sequential phases.

[43] Wendy Nelson Espeland and Mitchell L. Stevens, 'Commensuration as a social process', *Annual Review of Sociology*, 24 (1998), p. 315.

[44] Ibid., p. 316.

[45] Fioramonti, *How Numbers Rule the World*, p. 192.

[46] Peter Andreas and Kelly M. Greenhill, 'Introduction: the politics of numbers', in Peter Andreas and Kelly M. Greenhill (eds), *Sex, Drugs, and Body Counts: The Politics of Numbers in Global Crime and Conflict* (Ithaca, NY: Cornell University Press, 2010), p. 17.

numbers to generate information in a format that can be more easily and quickly assimilated by non-expert audiences, who might otherwise be overwhelmed by qualitative and contextual detail.

Commensuration requires fixed, stable, and universal categories. These are generated by means of reification, which refers to the translation of complex phenomena into observable and quantifiable conceptual categories that are presumed to be universally applicable irrespective of cultural or historical context. Reification effectively stabilises the meaning of complex and highly contested categories, such as democracy, freedom, and stability.[47] These reified categories in turn provide a foundation for different types of numerical assessment, the most notable of which are rankings, scales, and grades. Ranking consists of assigning individual units of analysis a position relative to their peers, such as country A being number one while country B is number six. Numerical scales, such as 1–10, produce the appearance of more precise and fine-grained measurement, with the performance of different actors being assigned a specific score out of a fixed total. In contrast, grades classify and group together multiple peers into defined qualitative bands, such as Free to Not Free, or Tier One to Tier Three. Grades are frequently represented using 'heat maps', with shades of green, yellow, and red being assigned to countries on a regional or global map based upon the specific grades they have been awarded. As benchmarking systems have evolved, the types of numerical assessments they generate have become ever more elaborate, but none of these assessments are possible without a foundation of fixed and unproblematic categories that create the appearance of certainty, coherence, and consistency.

Quantification and reification pave the way for symbolic judgments, in which the question of relative performance or value takes centre stage. Symbolic judgements on countries' relative performance are qualitatively different from what can be termed 'regulatory judgements', such as a determination that a government's actions constitute non-compliance with prescribed behaviour under the terms of an international agreement.[48] *Regulatory judgements* are more likely to involve direct and easily observable political consequences on target actors, whereas *symbolic judgements* are more likely to produce indirect political consequences through shaming processes, unfavourable comparisons with peers, and other forms of reputational damage. They may also generate reference points that are carried into other types of transnational practices, such as multilateral lending and development assistance, bilateral diplomatic relations, access to capital markets, or international programmes for intervention and policy reform.

Nearly all global benchmarks suffer from a 'dodgy data' problem. This problem can be particularly acute in cases where many different benchmarks are used in order to create composite benchmarks, resulting in a proliferation of data that frequently rests on very tenuous foundations.[49] Many benchmarkers are reluctant to make their methodology public, since this could complicate or undercut their market position or organisational credibility. Therefore, it often remains a mystery how specific

---

[47] Theodore M. Porter, *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life* (Princeton: Princeton University Press, 1995), p. 86.

[48] Beth A. Simmons, 'Compliance with international agreements', *Annual Review of Political Science*, 1 (1998), p. 77.

[49] Kevin E. Davis, Benedict Kingsbury, and Sally Engle Merry, 'Indicators as a technology of global governance', *Law and Society Review*, 46:1 (2012), p. 76.

conclusions were reached. This backstory routinely gets obscured once numbers are put into the public domain.

In many cases, the main political and institutional advantages associated with creating and disseminating benchmarks are political and organisational, rather than analytical. Global benchmarks have not only become relatively cheap and easy to produce and disseminate, they have also become increasingly popular among funders and donors eager to capture media attention. Most benchmarking does not involve years of expertise in the field, or contextual knowledge of local languages, customs, and social norms. Instead, all that is often required is the capacity to compile and process different forms of secondary data, which may simply involve aggregating and transposing information from one benchmark in order to create another.

The politics of global benchmarking

Global benchmarking typically relies on productive forms of indirect power to provoke reactions from target actors, with 'productive power' understood as the 'socially diffuse production of subjectivity in systems of meaning and signification'.[50] Much of the value of benchmarking, at least from a public relations or political activism standpoint, stems from the fact that benchmarks can play a key role in both stimulating and structuring political conversations regarding: (1) the dimensions, ramifications, and salience of a given set of issues; (2) how the performance of specific actors compares with that of their peers; and (3) how the performance of specific actors has changed with the passage of time. Benchmarking practices also tend to provoke politically motivated conversations around questions of methodology, whereby the credibility of particular measures is either impugned or defended depending on whether the results align with the political and economic agendas of the various actors involved.

While benchmarks purport to describe 'things as they are', this veneer of numerical representation and neutral comparison invariably conceals a range of political calculations, agendas, interests, and effects. Any overall assessment of 'good' or 'bad' performance requires a series of prior normative judgements regarding the types of activities, institutions, or categories that merit being subjected to benchmarking in the first place. At this juncture, it is essential to recognise that global benchmarking efforts almost invariably draw upon a common portfolio of normative values, assumptions, and agendas, such as liberal or neoliberal models of the rule of law, freedom of speech, democracy, human development, environmental protection, poverty alleviation, 'modern' statehood, and 'free' markets.

These normative commitments typically have a similar point of origin and influence, with Western experiences, assumptions, and paradigms exercising a disproportionate influence over the shape of international policy agendas and the articulation and definition of global problems. Moreover, Western states tend to populate the highest rankings across numerous benchmarks, with many non-Western states in turn receiving the lowest scores. While this is by no means a perfect relationship, since some non-Western states now feature amongst the 'high achievers', it is still possible to identify clear concentrations of Western and non-Western states at opposing ends of the spectrum across many global benchmarks.

---

[50] Barnett and Duvall, 'Power in international politics', p. 43.

	HDI*	Corruption	Freedom	Stability	Credit	Slavery	Business
High-performing European countries							
Norway	1	5	1 (free)	175	Aaa	150	9
Switzerland	3	7	1 (free)	174	Aaa	150	29
Netherlands	4	8	1 (free)	166	Aaa	139	28
Germany	6	12	1 (free)	165	Aaa	136	21
Denmark	10	1	1 (free)	176	Aaa	150	5
Ireland	11	21	1 (free)	170	Baa1	160	15
Sweden	12	3	1 (free)	177	Aaa	150	14
Iceland	13	12	1 (free)	171	Baa3	160	13
UK	14	14	1 (free)	161	Aa1	160	10
France	20	22	1 (free)	160	Aa1	139	38
Low-performing African countries							
Mozambique	178	119	3.5 (partly free)	50	B1	35	139
Guinea	179	150	5 (partly free)	12	Unrated	17	175
Burundi	180	157	5 (partly free)	21	Unrated	41	140
Burkina Faso	181	83	4 (partly free)	39	Unrated	26	154
Eritrea	182	160	7 (not free)	23	Unrated	15	184
Sierra Leone	183	119	3 (partly free)	35	Unrated	13	142
Chad	184	163	6.5 (not free)	6	Unrated	34	189
CAR**	185	144	7 (not free)	3	Unrated	31	188
Congo (DRC)	186	154	6 (not free)	4	B3	23	183
Niger	187	106	3.5 (partly free)	19	Unrated	28	176
N =	187	177	195	178		162	189
Metric	Ranking (lower better)	Ranking (lower better)	(lower better)	Ranking (higher better)	Grade (higher better)	Ranking (higher better)	Ranking (lower better)

Notes: *The table is composed of the 10 best-performing European and 10 worst-performing African states which feature across multiple benchmarks based on the 2014 Human Development Index (HDI). **Central African Republic.

Sources: 2014 Human Development Report; 2013 Corruption Perceptions Index; 2014 Freedom in the World Report; 2014 Fragile State Index; Moody's Investors Service (accessed 9 October 2014); 2013 Global Slavery Index; 2013 World Bank Doing Business Survey.

Table 1. *Comparing European and African countries across global benchmarks*

Table 1 serves to illustrate this underlying relationship by comparing the rankings recently assigned to ten 'high performing' European and ten 'low performing' African countries across a number of high-profile global benchmarks, including human development, corruption, freedom, state stability, credit, slavery, and business. The selection of ten countries is deliberate, since press releases and other materials that accompany the publication of benchmarks frequently concentrate attention on the top ten 'best' or 'worst' performers, who are specifically singled out for either condemnation or praise.

Each benchmark listed in this table is global in both scope and ambition. This global reach builds upon an underlying premise that there are certain values and criteria that can and should be treated as universal, irrespective of historical, political, or cultural differences. This type of universalism generates considerable controversy and norm contestation when expressed in other formats,[51] but benchmarks have proved to be an effective means of at least partially shielding normative arguments and agendas via appeals to models of neutral and technical assessment. Since the underlying normative commitments associated with an individual benchmark often closely align with those of other benchmarks in related domains, global benchmarking tends to have the cumulative effect of: (1) both reifying and generalising specific

[51] Kelley, *Monitoring Democracy*, p. 23.

models of governance, social organisation, and public policy; and (2) legitimating and promoting the recent histories and ongoing activities of Western states and, by extension, a variety of Western transnational actors.

In our opening discussion we divided benchmarking into three areas: quality of conduct, quality of design, and quality of outcomes. In the case of quality of outcomes, it is important to take into account a widespread tendency to assign singular responsibility for 'good' or 'bad' outcomes to the internal efforts of states and their peoples. When the United Kingdom receives a positive ranking, it is presumed to be the result of the internal efforts of the British state and its citizens, rather than as a consequence of interactions between Britain and other parts of the globe. Similarly, when Nigeria receives a negative ranking, it is tacitly presumed to be a result of the internal failings of the Nigerian state and society, rather than a consequence of external intrusions or structural conditions in the international system. This is highly problematic from an analytical standpoint, because the sources of 'good' or 'bad' performance tend to be far more diffuse than this model of responsibility suggests. There are many occasions when 'successful' states, along with numerous non-state actors, are at least partially responsible for the 'failures' of their peers. To give a stark example that illustrates this point: Iraq today scores poorly on a host of benchmarks, but how much of this is the responsibility of Iraqis?

This analytical slippage between outcomes and responsibility can be politically valuable for Western governments, populations, and corporations. Since high scores are widely presumed to be the result of individual efforts and achievements, global benchmarks frequently end up tacitly legitimating the wealth and privilege enjoyed by many actors in the West. Since low scores are widely presumed to be the result of internal failings and shortcomings, the impact of external actors and forces – most notably colonialism and imperialism – gets excluded from the political calculus.[52] This basic formula is in turn likely to provide further justification for particular forms of intervention and analysis, whereby Western actors can be represented as saviours and non-Western actors can be reduced to supplicants in need of paternalistic assistance. This formula obviously comes with a host of problems. In particular, no benchmark that assigns responsibility for outcomes that it is beyond the capacity of the 'responsible' party to address will be effective in bringing about change.[53]

These languages of legitimation and exculpation comprise one component of the larger politics of 'good' and 'bad' performance. In the case of the former, benchmarks tend to help reinforce established policies and organisational practices through the validating effects of favourable scores and superior rankings. This also extends to improvements in performance, where political leaders and other actors routinely claim credit when their countries and organisations have improved in global rankings, and may also seek to harness 'improvement' in order to attract interest from investors and aid agencies.[54] In these types of cases, benchmarks frequently become an

---

[52] This inward attribution of Western success is explored further in John Hobson, *The Eastern Origins of Western Civilisation* (Cambridge: Cambridge University Press, 2004); see also Shogo Suzuki, *Civilization and Empire: China and Japan's Encounter with European International Society* (Abingdon: Routledge, 2009); on the assessment of 'legitimacy' in terms of Western policy standards, see Lena Rethel, 'Whose legitimacy? Islamic finance and the global financial order', *Review of International Political Economy*, 18:5 (2011), pp. 75–98.

[53] See Ole Jacob Sending and Jon Harald Sanders Lie, 'The limits of global authority: How the World Bank benchmarks economies in Ethiopia and Malawi', *Review of International Studies*, 41:5 (2015), pp. 993–1010.

[54] Alexander Cooley, 'The emerging politics of international ranks and ratings: a framework for analysis', in Alexander Cooley and Jack Snyder (eds), *Ranking the World: Grading States as a Tool of Global Governance* (Cambridge: Cambridge University Press, 2015), p. 4.

instrument of status quo legitimation, and may be further invoked in order to deflect or dismiss calls for a different course of action.

The politics of 'bad' performance pull in a different direction, with negative or falling rankings providing an impetus for overhauling existing laws and policies, or at least providing political ammunition for critics of the status quo. Here, benchmarks can potentially prompt actors to 'alter their behaviour in reaction to being evaluated, observed, or measured'.[55] This can occur either *ex ante*, when actors anticipate future costs associated with a benchmarking exercise and seek to avoid the possibility of reputational damage, or *ex post*, when target actors observe and then respond to the costs associated with a specific result.[56] Unfavourable rankings in different global benchmarking regimes may result in either material sanctions (such as economic costs) or social sanctions (such as shaming or peer pressure via instruments such as a 'watch list' or a 'blacklist'), or a mix of both.[57]

There are many instances where a 'poor' result may have little or no immediate political effects; neither the material nor the social sanctions associated with benchmarking have consistent or predictable effects upon the behaviour of target actors. The imposition of material sanctions on 'pariah' states has often proved to be counterproductive for altering behaviour,[58] while those that have already gained pariah status are unlikely to be constrained by being further shamed and ostracised by the international community or through other social sanctions.[59] Nonetheless, when benchmarks gain sufficient prominence and credibility to provide a strong rationale for political action, they can exert a significant influence as a means to 'legitimate policy goals, the choice of target populations, and policy tools'.[60]

The degree of analytical and methodological rigour that underpins the construction of global benchmarking regimes cannot sufficiently explain why they have emerged as such a popular mode of transnational governance. A more compelling explanation is that the growth of global benchmarking reflects a dynamic 'benchmarking market'. This is tied to growing demand for benchmarks as a form of 'evidence' to enhance broader processes of governance, such as the effective allocation of official development assistance, the identification of internal security threats, enhancing accountability mechanisms in transnational governance, tracking standards of corporate behaviour, or monitoring national compliance with international policy regimes. There will therefore be occasions when 'the demand for numbers generates a supply'.[61] Yet while rank orderings of conduct, institutional design, and economic, social, and political outcomes may fulfil a functional need for

---

[55] Wendy Nelson Espeland and Michael Sauder, 'Rankings and reactivity: How public measures recreate social worlds', *American Journal of Sociology*, 113:1 (2007), p. 6.

[56] J. C. Sharman, 'The bark *is* the bite: International organizations and blacklisting', *Review of International Political Economy*, 16:4 (2009), pp. 573–96.

[57] J. C. Sharman, *Havens in a Storm: The Struggle for Global Tax Regulation* (Ithaca, NY: Cornell University Press, 2006), p. 104.

[58] Alexandra Homolar, 'Rebels without a conscience: the evolution of the rogue states narrative in US security policy', *European Journal of International Relations*, 14:4 (2010), pp. 705–27.

[59] Edward Weisband, 'Discursive multilateralism: Global benchmarks, shame, and learning in the ILO labor standards monitoring regime', *International Studies Quarterly*, 44:4 (2000), pp. 643–66.

[60] Anne Schneider and Helen Ingram, 'Social construction of target populations: Implications for politics and policy', *American Political Science Review*, 87:2 (1993), p. 339.

[61] Peter Reuter and Edwin M. Truman, *Chasing Dirty Money: The Fight Against Money Laundering* (Washington, DC: International Institute of Economics, 2004), p. 22, available at: {www.piie.com/publications/chapters_preview/381/2iie3705.pdf} accessed 20 June 2014.

existing processes of transnational governance, they also produce new power relations wielded by one group of actors over others.[62]

The practice of global benchmarking is a prime example of transnational governance that works via knowledge practices rooted in authoritative expertise in order to extend power over disparate objects and subjects.[63] However, benchmarking is distinct from other forms of expert authority commonly utilised by state institutions and international organisations, because of the opportunities it provides for non-state actors – whether civil society organisations or corporate agencies – to employ knowledge practices in an attempt to limit or alter how public authority is used. It is therefore important to unpack the practice of global benchmarking into different types to gain a more fine-grained understanding of how various forms of benchmarking, promulgated by different types of actors, intersect, overlap, and compete with each other across contemporary processes of transnational governance.

A typology of global benchmarking

In Table 2, we distinguish between four types of global benchmarking practices: (1) statecraft; (2) international governance; (3) private market governance; and (4) transnational advocacy.[64] This divides benchmarking practices into types based on the class of actor that is engaged in benchmarking, namely states, international organisations, profit-based private institutions, and non-profit private institutions. We use the public-private distinction as a 'category of analysis' to denote the different forms of accountability and capacities of various benchmarkers, rather than as a 'category of practice'.[65] While useful for heuristic purposes, these analytic divisions do not preclude the possibility that other actors may use one type of global benchmarking for a different purpose. Using this typology, we have compiled a Global Benchmarking Database consisting of 205 benchmarks (as of June 2015), which is available at: {www.warwick.ac.uk/globalbenchmarking/database}.

Type I benchmarking is a form of *statecraft*, whereby global benchmarks are produced by national government agencies such as ministries of finance and foreign affairs to extend state power internationally through the projection of particularistic values and standards of behaviour as universal. This may also legitimate the use of other foreign policy tools, such as sanctions and foreign aid, based on the conception of benchmark judgements as objective and neutral assessments of conduct, institutional design, or performance. Type II benchmarking is a form of *international governance*, which is undertaken by international organisations such as the World Bank, the International Monetary Fund (IMF), the Organisation for

---

[62] Tim Büthe, 'Beyond supply and demand: a politico-economic conceptual model', in Kevin E. Davis, Angelina Fisher, Benedict Kingsbury, and Sally Engle Merry (eds), *Governance by Indicators: Global Power Through Quantification* (Oxford: Oxford University Press, 2012), p. 51; see also Joel Quirk, 'The anti-slavery project: Linking the historical and the contemporary', *Human Rights Quarterly*, 28:3 (2006), p. 576.

[63] Wendy Larner and William Walters, 'Globalization as governmentality', *Alternatives*, 29:5 (2004), p. 496; see also André Broome and Leonard Seabrooke, 'Seeing like an international organisation', *New Political Economy*, 17:1 (2012), pp. 1–16.

[64] On the construction of typologies see Alexander L. George and Andrew Bennett, *Case Studies and Theory Development in the Social Sciences* (Cambridge: MIT Press, 2005), pp. 237–8.

[65] Stein Sundstøl Eriksen and Ole Jacob Sending, 'There is no global public: the idea of the public and the legitimation of governance', *International Theory*, 5:2 (2013), pp. 213–37.

Types of global benchmarking	Monitoring agents	Examples

Statecraft	National government agencies	Trafficking in Persons Report
		Millennium Challenge Account
		  Country Rankings
		World Military Expenditures and Arms
		  Transfers Rankings

International governance	International organisations and regional institutions	Human Development Index
		  Gender Empowerment Measure
		  World Governance Indicators
		  Country Policy and Institutional
		    Assessment

Private market governance	Profit-based institutions, including financial services and consultancy firms	Emerging Markets Bond Indices
		  International Country Risk Guide
		  Sovereign Credit Ratings
		  Supply chain benchmarking

Transnational advocacy	Civil society organizations, think tanks, media organisations, and academics	Corruption Perception Index
		  Fragile/Failed States Index
		  Index of Economic Freedom
		  Climate Change Performance Index

Table 2. *Four types of global benchmarking practices*

Economic Cooperation and Development, and the United Nations Development Programme; or by regional organisations, such as the European Union. This differs from benchmarking as statecraft because the practice of Type II benchmarking is usually under the control of international bureaucracies rather than national policymakers and is less directly geared towards the promotion of an individual state's national interests, although states often seek to use Type II benchmarks as instruments of statecraft.

Type III benchmarking is a form of *private market governance*, which is undertaken by profit-based institutions and is one of the oldest forms of benchmarking. This includes sovereign credit rating, which has its roots in the late nineteenth and early twentieth centuries,[66] and internal measures of performance and quality control used by large firms (self-benchmarking),[67] which has become increasingly significant as transnational corporations have spread their business activities worldwide through global production networks. Type IV benchmarking is either explicitly or implicitly geared towards *transnational advocacy* in particular issue areas, and is primarily conducted by civil society organisations and non-profit think tanks, but may also include work by individual academics or academic research centres. In some instances Type IV benchmarking involves collaboration between non-profit institutions and profit-based institutions, and in particular media organisations, such as the Index of Economic Freedom, which is produced by the Heritage Foundation and the *Wall Street Journal*. We further illustrate our typology of global benchmarking practices by briefly discussing a prominent example of each of type.

[66] Timothy J. Sinclair, *The New Masters of Capital: American Bond Rating Agencies and the Politics of Creditworthiness* (Ithaca, NY: Cornell University Press, 2005).
[67] Larner and Le Haron, 'Global benchmarking'.

Benchmarking as statecraft

Benchmarking as statecraft can be conceived as a form of 'soft power' in world politics.[68] A prominent example of Type I benchmarking is the *Trafficking in Persons Report*, which has been produced annually since 2001 by the US State Department's Office to Monitor and Combat Trafficking in Persons, and is officially described as 'the U.S. Government's principal diplomatic tool to engage foreign governments on human trafficking'.[69] This was established through the Victims of Trafficking and Violence Prevention Act, signed into law in 2000, which mandates that unilateral sanctions be applied by the US government on countries that are deemed not to meet minimum standards for the elimination of human trafficking, based on the *Report*. These sanctions can involve exclusion from US non-humanitarian and non-trade-related foreign aid, as well as US opposition to government requests for IMF or World Bank loan programmes.

The *Report* divides countries based on three different tiers. Tier One comprises countries whose governments are assessed as fully compliant with minimum standards for the elimination of human trafficking. In Tier Two are countries whose governments are not fully compliant but are assessed as making significant efforts to comply, with those deemed to face severe problems included in a separate category on the Tier Two Watch List. In Tier Three are countries whose governments are judged as non-compliant and not making sufficient efforts to comply with these minimum standards.[70] The data for the *Report* is based upon information from US embassies, government officials, non-governmental and international organisations, published reports, news articles, academic studies, research trips to every region of the world, and information submitted via email to report tips on human trafficking.[71]

In the decade and a half since it was established, the *Trafficking in Persons Report* has attracted substantial controversy and has been criticised for a lack of impartiality and political bias, with the US accused of acting as a 'global sheriff'.[72] States that have posed significant foreign policy problems for the US, such as Cuba and Venezuela, have typically received poorer rankings than otherwise broadly similar countries. Allies of the US with questionable records on human rights have historically received more positive assessments, although there has recently been a modest effort to correct this perception by shaming some US allies.[73] The US government has also recently found it necessary to include material on its own anti-trafficking efforts, following sustained criticism that their own record was notably absent from the reports. While all benchmarkers invariably start with specific agendas of their own, it is not uncommon for annual benchmarking

---

[68] Joseph S. Nye, *Soft Power: The Means to Success in World Politics* (New York: Public Affairs, 2004).
[69] {http://www.state.gov/j/tip/rls/tiprpt/} accessed 11 July 2014.
[70] Anne Gallagher, 'Trafficking in Persons Report (Review)', *Human Rights Quarterly*, 23:4 (2001), pp. 1136–7; Anne Gallagher, 'Improving the effectiveness of the international law of human trafficking: a vision for the future of the US Trafficking in Persons Reports', *Human Rights Review*, 12:3 (2011), pp. 381–400.
[71] {http://www.state.gov/j/tip/rls/tiprpt/2014/226645.htm} accessed 11 July 2014.
[72] Janie Chuang, 'The United States as global sheriff: Using unilateral sanctions to combat human trafficking', *Michigan Journal of International Law*, 27:2 (2006), pp. 437–94; Annie Bunting and Joel Quirk, 'Contemporary slavery as more than rhetorical strategy', in Annie Bunting and Joel Quirk, *The Invention of Contemporary Slavery: Studies in Rhetoric and Practice* (Vancouver: University of British Columbia Press, 2016).
[73] Gallagher, 'Improving the effectiveness of the international law of human trafficking', pp. 382–4.

exercises to evolve in unexpected ways, or to produce unpredictable findings and outcomes that complicate the original motivations for introducing the benchmark.

The *Report* is widely recognised as a prime example of the use of benchmarking as an exercise in statecraft that seeks to compel global action in accordance with the expectations and agenda of the US government.[74] Despite the numerous flaws that have been identified in these annual reports and associated policies responses,[75] recent research by Judith Kelley and Beth Simmons suggests that this example of Type I benchmarking has been highly consequential for the behaviour of (some) target actors. Kelley and Simmons conclude that 'states are sensitive to monitoring, respond faster to "harsher" grades, and react when their grade first drops below a socially significant threshold'.[76] Combating trafficking is a cause that comes with a host of practical problems and collateral damages, and it remains an open question whether the *Report* has helped or hurt in this respect. Nevertheless, this does not negate the larger point that this is a benchmark that has been globally influential.[77]

Benchmarking as international governance

The growth of benchmarking as international governance has gone hand in hand with the expansion of various forms of surveillance by international organisations of country performance over the last four decades.[78] The World Bank's Worldwide Governance Indicators (WGIs) is a useful illustrative example of Type II benchmarking. Starting from 1996 and covering over 200 countries, these indicators aim to measure governance performance around the world across six dimensions: (1) voice and accountability; (2) political stability and absence of violence; (3) governance effectiveness; (4) regulatory quality; (5) rule of law; and (6) control of corruption.[79] The data for the WGIs incorporates several hundred variables from 31 different data sources and is based on perceptions of governance quality drawn from public opinion and expert surveys, civil society organisations, profit-based information providers, and government agencies.

The conceptual validity, data accuracy, and substantive meaning of the WGI measures have been subject to strong criticism.[80] For example, to construct global benchmarks of governance quality, governance is defined as:

the traditions and institutions by which authority in a country is exercised. This includes (a) the process by which governments are selected, monitored, and replaced; (b) the capacity of

---

[74] Chuang, 'The United States as global sheriff', pp. 456–7.

[75] Ronald Weitzer, 'New directions in research on human trafficking', *The ANNALS of the American Academy of Political and Social Science*, 653:1 (2014), pp. 6–24; Thomas Steinfatt, 'Sex trafficking in Cambodia: Fabricated numbers versus empirical evidence', *Crime, Law, and Social Change,* 56:5 (2011), pp. 443–62.

[76] Kelley and Simmons, 'Politics by number', p. 68.

[77] See, for example, the articles in 'Beyond Trafficking and Slavery', *openDemocracy*, available at: {https://www.opendemocracy.net/beyondslavery} accessed 3 June 2015.

[78] See André Broome and Leonard Seabrooke, 'Seeing like the IMF: Institutional change in small open economies', *Review of International Political Economy*, 14:4 (2007), pp. 576–601; Liam Clegg, 'Our dream is a world full of poverty indicators: the US, the World Bank, and the power of numbers', *New Political Economy*, 15:4 (2010), pp. 473–92; Löwenheim, 'Examining the state'.

[79] Langbein and Knack, 'The worldwide governance indicators', pp. 369–70.

[80] Ibid., p. 351.

the government to effectively formulate and implement sound policies; and (c) the respect of citizens and the state for the institutions that govern economic and social interactions among them.[81]

Observers have highlighted the partial and biased view of governance quality that the definition used for the WGIs represents, emphasising in particular that the scale of aggregation involved in the production of the WGIs constitutes a trade-off between reliability and precision. Each of the data sources used to produce the WGIs suffers from specific quality problems. These problems are likely to be further complicated by aggregation processes, since the number and type of data sources differ both across countries and over time.[82]

   The WGIs implicitly assume a particular meaning of governance as a universally accepted standard. As M. A. Thomas points out, while most governments are likely to agree that the 'rule of law' is an important dimension of effective governance, for a liberal democracy this might be understood as 'a state constrained by rules' while an authoritarian dictatorship might understand this to mean 'citizen obedience to government edicts'. For these reasons, the WGIs have been criticised for not recognising that 'a governance indicator is a hypothesis about measurement *and* about the nature of governance'.[83] Nevertheless, as an example of Type II benchmarking the WGIs have resonated across a wide range of third parties, and have become particularly influential in decision-making processes over foreign aid allocations as a new form of policy conditionality.[84]

Benchmarking as private market governance

Benchmarking has become an increasingly prominent feature of national and transnational economic governance, especially in the aftermath of the global financial crisis when a large proportion of the pre-crisis ratings of financial assets produced by credit rating agencies were found to be inaccurate.[85] In particular, sovereign credit ratings represent one of the most controversial examples of benchmarking as private market governance. Credit ratings are evaluations of a debtor's ability to repay a loan and the probability of default. As a form of Type III benchmarking, sovereign credit ratings by the three major ratings agencies – Moody's Investors Service, Standard and Poor's, and Fitch Ratings – impact upon governments' fiscal autonomy and the terms on which they can raise public debt. The symbolic judgements made by private ratings agencies affect the creditworthiness of national and local governments, firms, banks and other private companies, and in theory function to reduce uncertainty and information asymmetry problems for investors. Like other types of benchmarking,

[81] Aart Kraay, Daniel Kaufmann, and Massimo Mastruzzi, 'The worldwide governance indicators: Methodology and analytical issues', World Bank Policy Research Working Paper No. 5430 (2010), p. 4, emphasis in original, available at: {http://elibrary.worldbank.org/doi/pdf/10.1596/1813-9450-5430} accessed 15 July 2014.
[82] Steven van de Walle, 'The state of the world's bureaucracies', *Journal of Comparative Policy Analysis: Research and Practice*, 8:4 (2006), pp. 439–40.
[83] M. A. Thomas, 'What do the worldwide governance indicators measure?', *European Journal of Development Research*, 22:1 (2010), p. 50, emphasis added.
[84] Thomas, 'What do the worldwide governance indicators measure?', p. 32.
[85] Rawi Abdelal and Mark Blyth, 'Just who put you in charge? We did', in Alexander Cooley and Jack Snyder (eds), *Ranking the World: Grading States as a Tool of Global Governance* (Cambridge: Cambridge University Press, 2015), p. 46.

however, credit ratings 'not only provide information but help construct the context in which corporations and public bodies make decisions'.[86]

Standard and Poor's utilises a mix of qualitative and quantitative measures in the five factors that constitute its sovereign ratings. These include: (1) a 'political score', which focuses on the quality of political and policymaking institutions, and external risks; (2) an 'economic score', which incorporates the degree of economic diversity, income levels, and growth prospects; (3) an 'external score', based on the international status of a country's currency, external liquidity, and foreign debt levels; (4) a 'fiscal score', based on assessments of the sustainability of budget deficits and public debt burden; and (5) a 'monetary score', which is based on inflation rates, the degree of flexibility in monetary policy, and the depth of domestic financial markets. Standard and Poor's credit analysts assign a score for each of these five factors ranging from one (strongest) to six (weakest).[87] Once a sovereign credit rating is officially assigned to a country, ratings are then monitored on an ongoing basis and reviewed at least once a year.

Ratings issued by the three major agencies constitute a rank ordering of credit risk. Long-term ratings, for example, are distinguished between different ranks of 'investment grade' ratings (ranging from the top AAA rating, issued by Standard and Poor's and Fitch Ratings, to the BBB rating) and 'non-investment grade' or 'speculative grade' ratings (BB ratings and below.) Ratings below investment grade are considered to have a moderate to high credit risk of non-repayment. The power of the symbolic judgements rating agencies issue comes primarily from their role as 'reputational intermediaries'. This is based on their public image as independent, authoritative actors that are capable of making accurate expert assessments of creditworthiness, despite this image being subjected to stringent criticisms in recent years.[88] Moreover, credit rating agencies also remain highly controversial because their benchmarking activities help to reify and consolidate international norms of 'proper fiscal conduct', which shapes perceptions about what constitutes 'normal' economic behaviour by governments.[89]

Benchmarking as transnational advocacy

The use of global benchmarking by NGOs engaged in transnational advocacy has risen dramatically in recent years. Among other reasons for this growing trend, many funders value the capacity of benchmarks and indicators to provide 'clear signals' that can be used as proxies for measuring the relative success of political campaigns and policy interventions.[90] In June 2014, for example, the US think-tank Global Fund for Peace launched the tenth edition of their 'Fragile States Index', which is an example of Type IV benchmarking as transnational advocacy. Thanks to an established partnership with *Foreign Policy* magazine, the headline findings of the index had

[86] Fioramonti, *How Numbers Rule the World*, pp. 60–1.
[87] Standard and Poor's, *Sovereign Government Rating Methodology and Assumptions* (New York: Standard and Poor's, 2011) available at: {www.standardandpoors.com/ratings/articles/en/us/?articleType=PDF&assetID=124531655404} accessed 15 July 2014.
[88] Sinclair, *The New Masters of Capital*, p. 176.
[89] Bartholomew Paudyn, 'Credit rating agencies and the sovereign debt crisis: Performing the politics of creditworthiness through risk and uncertainty', *Review of International Political Economy*, 20:4 (2013), pp. 799–800.
[90] Sarah Bush, *The Taming of Democracy Assistance: Why Democracy Promotion Does Not Confront Dictators* (Cambridge: Cambridge University Press, 2015), p. 14.

been precirculated. Among the highlights of the 2014 report was the news that 'after six years in the number one position [as the world's worst 'failed state'] Somalia has finally been overtaken, leaving South Sudan as the most fragile state in the world'.[91]

The publication of the Fragile States Index (FSI) is intended as a form of transnational advocacy. With the stated goal of encouraging 'discussion, advocacy and action on the underlying conditions that could create conflict and do threaten human security and economic development',[92] the FSI ranks 178 states on the basis of their 'levels of stability and the pressures they face'.[93] This high-profile exercise in global benchmarking involves assigning states a numerical value (1–120) based on their relative vulnerability to 'state failure', and grouping states into different categories ranging from 'high alert' to 'very sustainable', with shades of red, yellow, and green used to highlight their relative status. These findings are generated using a patented 'Conflict Assessment Software Tool', which applies 'sophisticated search parameters and algorithms' to separate relevant from irrelevant data in the analysis of millions of documents each year.[94] This software tool integrates data from 'twelve primary social, economic and political indicators' (which include over 100 sub-indicators), with themes such as 'state legitimacy', 'factionalized elites', 'group grievance', 'security apparatus', and 'poverty and economic decline', yet the specific sources of raw data associated with these indicators have not been revealed.

The 'failed state' concept has been subject to sustained critique. Some of these criticisms were partly acknowledged by the decision by the Fund for Peace to rename the benchmark in 2014 as the 'Fragile States Index' to replace the former title of the 'Failed States Index'. The concept of 'failed states' has been repeatedly denounced for – amongst other things – lumping together states with very different histories and problems, for normalising a particular vision of 'modern' statehood and 'state-building', for directing responsibility for 'failure' inwards, rather than looking at external actors, and for being too closely aligned with US foreign policy goals.[95] This highlights the ease with which other actors might use one type of global benchmarking for different purposes, in order to pursue a larger set of interests and agendas.

Global benchmarking and third party users

The most influential users of each of these four types of global benchmarking are often third parties. These can be either public or private actors who may not be the formal target of a particular benchmarking exercise, but who incorporate benchmark scores produced by other actors into their decision-making processes and

advocacy efforts.[96] This use by third parties can greatly expand the political traction of benchmarks by multiplying the reputational costs or benefits associated with specific rankings, and intensifying competitive pressures to improve poor performance.

Freedom House, for example, does not carry much independent weight as an organisation. As an advocacy-oriented NGO, it is unable to use material incentives to induce compliance; its symbolic judgements on country performance do not imply the same potential for direct consequences as negative country reports issued by international organisations such as the World Bank or the IMF, and its claim to expert authority has been subject to strong reputational challenges on the basis of methodological weaknesses.[97] Nevertheless, its 'Freedom in the World' benchmark has acquired substantial weight owing to its alignment with the interests and agendas of third parties such as the US government and various international organisations, which greatly magnifies its audience and influence.

In such cases, the scientific expertise or institutional status of the benchmarker may be less consequential than what other parties do with their benchmark once it is produced. This underscores the need for more nuanced analyses of how global benchmarking links up with other transnational practices, as well as how benchmarks can potentially lead to unintended consequences. To gain a richer understanding of political effects of global benchmarking, it is therefore necessary to take into account: (1) the status and history of the specific organisation or individual that has produced a given benchmark; (2) the internal mechanics of how a given benchmark is produced; (3) the distinctive characteristics and political and economic profile of the specific issue being benchmarked; and (4) the authority and credibility that third party users can invest in benchmarks when they align with other political interests and agendas.

Conclusion

Global benchmarks are inspired by frequently overlapping normative values and agendas, which are then translated into ratings, rankings, and measurements for a given category of conduct, institutional design, or outcome. They are designed to promote distinctive forms of transnational behaviour and transnational organisation by enabling symbolic judgements of performance that are expressed through numerical values. These numerical values create information shortcuts that facilitate non-expert comparisons of 'good' or 'bad' performance by radically simplifying both context and complexity. Thanks in part to the popular and political appeal of numbers, benchmarks have emerged as a key practice for both promoting and codifying many different agendas and interests, and for either legitimating or challenging a diverse range of global actors and transnational activities. While it is in the interests of benchmarkers to rhetorically appeal to models of neutral, methodical, and technocratic assessment, their activities and outputs will always be inherently political.

Global benchmarking raises a number of critical questions for IR scholars and future IR research. Benchmarking is now routinely deployed as a tool of governance

---

[96] Sabino Cassese and Lorenzo Casini, 'Public regulation of global indicators', in Kevin E. Davis, Angelina Fisher, Benedict Kingsbury, and Sally Engle Merry (eds), *Governance by Indicators: Global Power through Quantification* (Oxford: Oxford University Press, 2012), pp. 468–9.

[97] Homolar, 'Benchmarking human security'; see also Christopher G. Bradley, 'International organizations and the production of indicators: the case of Freedom House', in Sally Engle Merry, Kevin E. Davis, and Benedict Kingsbury (eds), *The Quiet Power of Indicators: Measuring Governance, Corruption, and Rule of Law* (Cambridge: Cambridge University Press, 2015), pp. 27–74.

and knowledge production across a wide range of transnational policy arenas, and there are important differences between the four main types of benchmarking that we have identified and the political impact they can have in world politics. Moreover, the growing use of global benchmarks as a tool for constructing (at least the appearance of) authoritative expertise, and for extending public and private authority over distant entities, has increased the need to connect theories of how power operates indirectly in the international realm to explanations of how and why such efforts are – *or are not* – successful at achieving their intended ends.

Accordingly, we suggest that a new research agenda for the study of global benchmarking should take on board the following lines of inquiry: How are benchmarking practices defended and legitimated, and among which audiences and in the context of which markets for activism and advocacy? Why do specific benchmarks gain traction, both among target actors and third parties, while others fail to secure an audience? Why and how does a specific benchmark have an impact in one country while remaining inconsequential in others that share broadly similar features? What types of activities and effects do global benchmarks tend to obscure or conceal? How can we better understand the long-term consequences and costs associated with benchmarking in relation to contested issues such as responsibility, accountability, and private governance? How does the practice of global benchmarking revitalise or deepen existing IR literatures relating to transnational advocacy networks, global governance and governmentality, transnational actors, rational design and cooperation, and the politics of expertise? The various contributions to this Special Issue should not only help scholars to better understand the politics of numbers and normative agendas in global benchmarking, they should also help us to in turn ask better questions about how and where the practice of global benchmarking fits within broader patterns, processes, and theories of International Relations.