Political leaders care about their reputations, especially personal reputations relative to their peers. Entrepreneurial actors have seized on this fact as an opportunity for policy influence. They take advantage of the competitive pressures of globalization, the increasing accountability demands flowing from democratization, and the transformative capacities of new information technologies. This all to nudge global governance in important ways by deploying what we call Global Performance Indicators (GPIs) – regularized public assessments that rate, rank, and categorize state policies, qualities, and/or performance. Such GPIs aim to (re)define global standards, contrive competition, and provoke action according to specific performance criteria. They exemplify how states, intergovernmental organizations (IGOs), non-governmental organizations (NGOs), and think tanks seek to use modern information politics to invoke reputational concerns and influence policy through the pressure of comparison.

Performance assessments and associated indicators permeate human activity, internationally and nationally, across public and private spheres. Some, such as product ratings in Consumer Reports or national credit ratings by Standard and Poor’s (S&P), have been around for decades; many are intended to inform private decision-making. More recently, agenda-motivated indicators, such as the country ratings of Freedom in the World by Freedom House or the Ease of Doing Business Index by the World Bank, have appeared.† Such indicators overtly and purposively seek to engage actors at the highest levels of government and influence their policies and governance practices.

2 Judith G. Kelley and Beth A. Simmons

The rising use of indicators has prompted debates about their construct validity, use, and misuse.\(^2\) Scholarship has flourished about the methodology, reliability, and fairness for a broad range of indicators and has problematized quantification.\(^3\) We build on a range of insights – from sociology to anthropology to political science – and we focus on intentional efforts to use comparative information to influence policy and governance. Why do states and other important actors care about ratings and rankings? Is effective use of GPIs confined to already powerful actors such as the United States, or can GPIs contribute to a diffusion of authority? And if a GPI is effective, does it work primarily through material leverage, or are other mechanisms, such as reputational or social pressure, at work as well?

This volume expands our ability to answer these questions by examining the effects of eleven different GPIs spanning issues such as terrorism, corruption, democracy, business regulation, aid transparency, education, global governance, and labor rights. The chapters present evidence that these GPIs influence discourse, standards, and measurement, but they also examine whether and how GPIs change actions. Do they reverberate in domestic politics, provoke bureaucracies, shape policy agendas, and influence outputs? Do they influence the reactions and responses of donors and market actors? If they alter priorities, do they have unintended consequences as well? This volume aims to speak directly to how GPIs might alter politics, policies, and, more speculatively, even broader authority structures.

GLOBAL PERFORMANCE INDICATORS: DEFINITION, SCOPE, AND CONTEXT

Global performance indicators are a named collection of rank-ordered data that purports to represent the past or projected performance of different units.\(^4\) While thousands of basic indicators such as Gross Domestic Product (GDP) exist to facilitate research or policymaking, and while all such quantification is normative and contestable,\(^5\) this volume focuses on the consequences of overtly strategic state rating and ranking systems that package and deploy information intentionally for policy advocacy and implementation. Such GPIs assemble selected data – facts and statistics collected together for reference or analysis – to craft “new” information designed and publicized to complement advocacy for specific ends. We focus on GPIs that, in their purest form, meet all the following criteria. They are:

\(^2\) Malito, Umbach, and Bhuta 2018.
\(^3\) Arndt and Oman 2006; Ginsburg et al. 2005; Høyland, Moene, and Willumsen 2012; Van de Walle 2006. The World Bank’s Governance Indicators are a potent example; see Apaza 2009. For a critique of their construct validity see Thomas 2010.
\(^4\) Davis et al. 2012; Davis, Kingsbury, and Merry 2012; Davis, Merry, and Kingsbury 2015; Merry and Conley 2011.
\(^5\) Fioramonti 2013.
Introduction

- Public and easily available;
- Regular and published on a predictable schedule;
- Purposive, explicitly normative, policy focused;
- Deployed to influence state-level outcomes; and
- Comparative of the performance of multiple states within a region or more broadly.

Global performance indicators can take several forms. Indices or indicators use numbers or grades to rate or rank state performance, compressing enormous variance into a simplified scale. Categorical assessments use ordinal categories to produce (un)flattering peer groups, while blacklists and watchlists draw stark distinctions between compliers and offenders. Hybrid systems are also common, as are accompanying country narratives and policy recommendations.

Basic indicators are hardly new – sovereign credit ratings first appeared in 1916 – but nearly all the GPIs that fit our definition were created after 1990. To demonstrate the trend, we systematically gathered a database based on secondary literature, extensive searches of newspapers, magazines, and the wider web, using terms such as ranking, rating, index, blacklist, and watchlist. Each entry was coded, among other things, on launch year, regularity, purpose, creator type, country of origin, issue area(s), and the type of comparison used. To be included, a GPI had to have been created no later than 2015. This search produced 282 entries. Extensive culling left 159 entries that met our criteria, and of these, 133 were still active as of 2018, meaning they had been updated in 2015 or later. These 159 were included in the source data for Figure 1.1.

Global performance indicator growth has been nearly exponential. While it took almost three decades to approach twenty GPIs in the late 1990s, by the next decade the number had roughly quadrupled, and then in the next fifteen years it more than tripled. This proliferation responded to a growing demand for policy-relevant performance data, spurred by a consensus that many past development reforms and governance structures were failing, and facilitated by the fact that information was becoming ever easier to collect, process, and disseminate.

---

6 Broome and Quirk 2015; Büthe 2012; Davis et al. 2012; Hansen 2012; Karpik 2010.
7 Löwenheim 2008.
8 Of the 289, 29 were never reissued, so they failed the regularity criteria. Another 28 were aimed at investors, so they failed the policy purpose. Thirty had no explicit intent stated, so while they might qualify, they fell in a gray zone and for clarity of criteria, we noted them separately. Another 31 were just data (and clearly, we could have found hundreds more that would fit this category). A few others fell in other categories that didn’t fit.
9 The data collection concluded in July 2018.
10 Arndt 2008; Arndt and Oman 2006. Governance measures are the major theme in the Malito et al. 2018 essay collection.
Judith G. Kelley and Beth A. Simmons

As both the demand and supply of indicators grew, various actors began to realize that they could deploy GPIs as policy tools. A host of new GPIs, ranging from Transparency International’s *Corruption Perceptions Index* to the Organisation for Economic Co-operation and Development’s (OECD’s) *Program for International Student Assessment* (PISA) rankings began to flourish. Their labels became more normative and vivid, their content ever more actionable, and their messages increasingly displayed to deliver visual impact. Recent GPIs are relentlessly comparative, suggesting an intention to pressure, shame, or provoke competition among states. Competitive prompts are clear in many GPIs’ publicity messages. For example, when releasing the latest PISA rankings, the OECD announced: “Singapore tops latest OECD PISA global education survey.”11 Thus, GPIs are increasingly deployed to stimulate state competition and shape policy agendas; that is, as “technologies of power.”12

---


12 Hansen 2012.
Introduction

A THEORY OF GPI POWER

Why should governments care about a simple ranking or rating, especially if it merely reformats pre-existing information? We theorize that the power of GPIs is a function of their ability to engage reputations. Ratings and rankings stoke politicians’ and bureaucrats’ reputational concerns by framing, establishing “standards,” and repeatedly engendering public comparisons. The packaging and intentional deployment of comparative information is, thus, an inherently social process that stimulates reputational concerns with how others view the performance of the state elites’ competence, effectiveness, and status. These concerns can operate through multiple pathways, depending on local values and institutions.

Sources of GPI Power: Information, Contestation, and the Power to Frame

The social process of GPI influence begins with their ability to frame issues. The GPI promulgators seek to influence targets by designing an attention-grabbing form of political communication. By naming an issue, coining a vocabulary for describing it, and creating categories for its assessment, promulgators hope to affect discourse and, ultimately, policy.13 This influence is akin to what Davis and colleagues call “knowledge effects”14 and what Pierre Bourdieu refers to as “symbolic power” – or the power to name and categorize.15 In some cases, the process begins by collecting new data; in others, existing data are presented in ways that prioritize new objectives. Such efforts shape how the public, organizations, and policymakers think about “what constitutes legitimate social practice.”16

The GPIs are both products of, and inputs into, active value contestation. Nowhere is this clearer than in the struggle to define, measure, and rate “development” over recent years.17 The United Nation’s (UN) Human Development Index was created in 1990, adding lifespan, education, and low fertility to traditional GDP per capita measures to reflect new concerns with the quality of life. A decade later (2000), the new UN Gender Equality Index looked at disparities between men’s and women’s quality of life as the central conceptual issue.18 The Fraser Institute contested both of these trends with its rival Index of Human Progress (2001) that restored weight to high-income

13 Scheufele and Tewksbury 2007.
14 Davis, Merry, and Kingsbury 2015.
16 Hansen 2012.
17 For the history of the Human Development Index, see Stanton 2007. For the World Bank’s Governance Indicators see Kaufmann, Kraay, and Zoido 1999.
levels and reflected the institute’s interest in supporting competitive markets. In 2006, the Sustainable Society Foundation, an NGO headquartered in the Netherlands, touted its eponymously named index to contest the notion that unbounded growth is desirable. In short, creators of GPIs frame successful performance in preferred directions, often crowding the market with GPIs that jostle for agenda-setting visibility.

Figure 1.1 indicates evidence of contestation across multiple GPI sectors. It shows the gradual uptake of GPIs in various issue areas. A lack of consensus may be one motivation for creating new GPIs and may partly explain the variation in density across issues. For example, while there is likely a greater demand for information about economic and financial performance, which has led to a greater supply by actors keen to dominate such visible issues, it is also likely that the lack of consensus around what constitutes favorable policies and ideal outcomes has contributed to the proliferation. Although we excluded indicators created simply to inform investors, over a third of the GPIs meeting our definition (36%) focused on economic issues with several alternative measures of prosperity and well-being to contest the traditional per capita GDP. Many GPIs are also concerned with various education or social issues (29%), governance (20%), development (20%), and the environment (17%), with several clusters covering hybrids of these issues as well. Many such hybrids represent efforts to contest narrow measures and advocate for broader conceptualizations of policy priorities. This is one reason the classification of GPIs in issue areas results in several that fall in multiple categories. Many are meant precisely to pull traditional conceptions, such as growth, in an unconventional direction, for example, toward justice.

Social Knowledge and Authority

A key source of GPI power is the credibility and authority of its creator. What makes some GPI creators more authoritative than others? The social psychology literature suggests that one source of legitimate authority is trust, which develops out of a perception that an actor is fair, knowledgeable, and/or competent. This may explain why many GPI creators attempt to

---

10 See “Sustainable Society Index – Your Compass to Sustainability,” retrieved from www.ssfindex.com/
11 Miller 2005.
12 On private authority in international relations, see Hall and Biersteker 2002.
13 Espeland and Sauder 2007; Rieh 2002; Wilson 1983. Well-respected raters have accumulative trust over decades, so people may act on this “expert” information, even when it erroneous; see Simonsohn 2011. On the importance of legitimacy for the effectiveness of international organizations see Buchanan and Keohane 2006; Zürn 2004.
Introduction

bolster their credibility by transparently describing and defending their index’s methodology. Just as important, actors who themselves are thought to exemplify specific norms are more likely to be viewed as authoritative in judging them, and to have “normative power,” which Manners defines as the ability to shape conceptions of what is “normal.”

The GPI creators also gain authority based on their (assumed) competence and expertise, which may be inferred from their tangible power and wealth. The assumption that global competence in governance, economic policy, and human rights, for example, resides in the Western democracies, while contestable, has reinforced their “authority” to rate others in these areas. This may explain why, according to our data, organizations headquartered in the Global North create about 98 percent of the GPIs in existence today.

Social knowledge and authority are also supported by the tangible and intangible advantages of network centrality. Non-state actors with access to decision makers are better able to mine information and deploy it to assert authority. States also have important network resources useful in developing and deploying GPIs. For example, the US Department of State, which creates the Trafficking in Persons Report and grades country performance, uses its embassies and NGO contacts around the world to tap into many sources of information. Actors centrally located in a social and political network are better able to set agendas and have an impact on information flows, which facilitates both data collection and GPI dissemination.

Both states and non-state actors can accrue social authority for different reasons. States have advantages that flow from their pre-existing organization, broad resource base, and, in some cases, their political legitimacy. In contrast, non-state actors’ authority rests on independence from powerful political actors. An interesting case that Honig and Weaver explore is the Aid Transparency Index. They show that a small meagerly resourced NGO used its political independence and first-mover advantage to establish the gold standard for funding transparency. Political independence is plausibly the strength of several GPIs in our database; about half are created by NGOs and the rest are invented by universities, private actors, or a combination of actors. The United States is the only country that has created GPIs that target other states, which is only about 6 percent of the total. Overall, GPIs are largely not controlled by individual states or IGOs, although those created by the major IGOs and the United States attract more attention and, on average, may be more influential.
Judgment and Comparison: Provoking Status Concerns and Competition

Global performance indicators are consequential because they engage the reputation of the people or entities being assessed. A reputation refers to a widespread belief that a person or an organization can reliably be characterized in a particular way. A reputation and reputation are related, but status is explicitly comparative: it refers to the relative social or professional standing of someone or something in a formal or informal social hierarchy. Both reputation and status are quintessentially social constructs; they are granted or accorded only by a social community. The GPIs confer status, which cannot be credibly claimed or manufactured unilaterally by the target. What makes this form of information deployment potentially powerful is its ability to broadly affect social beliefs about successful states and appropriate policies. It is deployed precisely as a form of social pressure on targets to conform – or suffer the reputational consequences, real or imagined.

Exactly why people value their status and reputation or that of their country is hotly debated. States and their elites likely seek to cultivate reputations as a form of strategic capital to improve their payoffs. Some care about their reputations for reasons having to do with their social role or identity. The massive literature on shaming, for example, is built on the premise that officials want to avoid damaging their personal, professional, or organizational reputations. Bureaucrats, politicians, and their public also care about their own or their country’s relative standing, which can be hard to establish in a thinly institutionalized international environment. Status can be parlayed into material benefits and/or be an end in itself. International relations research accords with similar theories in economics, psychology, and sociology in accepting that people are as motivated by status as by power or material goods, with complex consequences for attitudes and behavior.

32 On the reputational impact of comparisons, see Kelley 2017, Chapter 2.
33 Masaki and Parks 2019 find that policy makers rate comparative assessments as more effective than single country assessments.
35 Tingley and Walter 2011.
36 Dafoe, Renshon, and Huth 2014.
37 On social pressures, see Checkel 2001; Johnston 2001. On shaming by states or IGOs, see Joachim, Reinalda, and Verbeek 2008; Lebovic and Voeten 2006. By non-governmental actors, see Hendrix and Wong 2013; Murdie and Davis 2012; Risse 1999.
38 Rosen 2005; Wohlforth 2009.
39 Dafoe, Renshon, and Huth 2014.
40 Rege 2008. Some scholarship even suggests that status can propel costly policy choices, even war, see Dafoe, Renshon, and Huth 2014.
41 Harsanyi 1966.
42 Blader and Chen 2012.
43 Ridgeway 2013.
Introduction

The GPI creators seek to engage this concern with status and reputation by leveraging comparative information among peers. Explicit comparisons create contexts in which judgments are formed and identities are established and reinforced. They foster commensuration, or the comparison of different entities according to a common metric, as a way of making highly simplified sense of the world. Labels like “rule of law,” “freedom,” “sustainable development,” “peacefulness,” or “political risk” imply that very different entities are, for the purposes of one concept, comparable. Numbers facilitate making standards through the simplest possible means – averaging.

The GPIs’ recurrent nature incentivizes governments to look to the next iteration of the index and worry about the reputational consequences of their next rating or ranking. The media is particularly fond of reporting relative rankings. In numerous interviews, GPI creators frankly acknowledged that they created such indexes precisely to attract media attention. About two-thirds of the active GPIs we could identify employ explicit top-to-bottom ranking systems, and over a third create clear normative categories or performance tiers, usually in addition to a ranking or rating. These features render GPIs a potent tool for producing social control through the pressure of comparative information.

Pathways to Responses: Domestic Politics, Bureaucratic Responses, and Transnational Pressures

This book explores how comparative rankings influence policy choices through multiple mechanisms. Figure 1.2 illustrates how GPIs reverberate through different levels of politics: domestic, elite, and transnational, and underscores the cyclical nature of iterative assessments.

First, GPIs can provide information that reverberates in domestic politics, especially when amplified by the popular or social media, or circulated in civil society. Important domestic audiences include local advocacy organizations, local
businesses, and even popular opinion generally. These groups use the comparative information from GPIs to demand better policy, especially when their country apparently lags behind. The underlying information need not be wholly new, but the assessment of relative performance is. The salience of the new information becomes an opportunity to mobilize in several ways: electorally through the ballot box, by publicizing critiques of performance in the media, engaging in traditional forms of lobbying, or by protesting or engaging in collective action. In responsive political systems, these demands might elicit policy change. At a minimum, officials will have the incentive to claim they are addressing the issue. However, where institutions repress public input and suppress political demands, governments may respond not with reform but by denigrating the GPI or its creator.\footnote{See Kelley 2017, Chapter 4.}

Second, many GPIs involve intensive consultations with government bureaucrats during which officials take “expert” advice directly into account in their policymaking.\footnote{Masaki and Parks 2019 find that GPIs on which bureaucrats are consulted are among the most salient to government officials. On the importance of bureaucrats’ reputations, see Chwieroth 2013; Nelson 2017.} Such monitoring signals the value that the monitor and other actors attach to specific tasks,\footnote{Larson and Callahan 1990.} and can influence targets to internalize monitoring regimes and eventually to self-regulate – a process that may mistakenly lead to underestimating the monitoring effects.\footnote{Self-regulation has been found effective in the context of health and safety monitoring by OSHA, see Johnson 2018.} Referred to as the “Hawthorne effect,” individuals in experimental settings have been found to rearrange their priorities to meet external expectations when they are aware of being observed.\footnote{Adair 1984.} This form of reactivity – the tendency for people to change their behavior in response to being evaluated – may explain the effect, for example, of \textit{US News & World Report’s} best colleges and best grad schools...