PALGRAVE HANDBOOKS

THE PALGRAVE HANDBOOK OF INDICATORS IN GLOBAL GOVERNANCE

Edited by Debora Valentina Malito, Gaby Umbach and Nehal Bhuta

The Palgrave Handbook of Indicators in Global Governance

"Knowledge was an essential part of the making of the modern State and we judge governments by measuring and ranking them. Indicators—their collection, application, use, and comparison— are hence an essential part of global governance. But how to measure and compare the worldwide rule of law, integrity, good governance, stateness, institutional quality, government performance, sustainable development, or human rights? This groundbreaking, multiperspective and transdisciplinary book provides all that one needs to explore this new area."

Sabino Cassese, *Professor of Global Law, LUISS Guido Carli University, former judge of the Italian Constitutional Court and Emeritus Professor at the Scuola Normale Superiore of Pisa*

"Malito, Umbach and Bhuta have put together a meticulous account of global governance indicators. The handbook takes stock of theory and practice of measurement, highlighting positive contributions and problematic areas for a wide range of governance aspects. This is an important contribution to the debate, useful to academics in many disciplines and policymakers alike."

Antonio Savoia, *Lecturer in Development Economics, University of Manchester*

Debora Valentina Malito • Gaby Umbach
Nehal Bhuta
Editors

The Palgrave Handbook of Indicators in Global Governance

palgrave
macmillan

Editors
Debora Valentina Malito
Robert Schuman Centre for Advanced Studies
European University Institute
San Domenico di Fiesole, Italy

Gaby Umbach
Robert Schuman Centre for Advanced Studies
European University Institute
San Domenico di Fiesole, Italy

Nehal Bhuta
Law Department
European University Institute
Firenze, Italy

ISBN 978-3-319-62706-9 ISBN 978-3-319-62707-6 (eBook)
https://doi.org/10.1007/978-3-319-62707-6

Library of Congress Control Number: 2017956093

Cover Image: © FL collection / Alamy Stock Photo
Cover Design: Tom Howey

Printed on acid-free paper

This Palgrave Macmillan imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

This Handbook is the outcome of the "Global Governance by Indicators" project, convened by the Global Governance Programme (GGP) of the Robert Schuman Centre for Advanced Studies, at the European University Institute (EUI). This project started with three workshops on measuring corruption, governance, and sustainability held at the EUI, in Florence, between 2013 and 2014. The workshops served as a platform for exchanges between scholars and practitioners to promote mutual understanding and learning about the role of indicators in global governance. The idea of writing this Handbook originated in a conversation we had about the heterogeneity, complexity, discordant, and contrasting orientations in the production and use of indicators. While observing the extraordinary proliferation of metrics, we were puzzled by the diffusion of scepticism about the power of indicators in informing and steering policy debates, as they were perceived as not actionable enough and their power as highly contextual. Such heterogeneity led us to collect this multi-perspective and trans-disciplinary range of contributions.

This Handbook had a long gestation, and as consequences, many debts are owed. We are grateful to the GGP at the EUI, for funding the "Global Governance by Indicators" project that made this work possible. Special thanks are also owed to the staff of the GGP for supporting the project as well as for managing the organisation of three workshops that led to this Handbook.

The workshops and chapters presented benefited from many valuable comments we received from colleagues and workshop participants. Particular thanks hence go to Mikai Akech, Matthias Brückner, Sabino Cassese, Mary Crane-Charef, Cristina Dallara, Kevin Davis, Luis de Sousa, Borbola Garai, Gustavo García, Julien Desmedt, Bernard Hoekman, David Hulme, Roberto

Martínez B. Kukutschka, Martina Kühner, Alina Mungiu-Pippidi, Carlos Closa Montero, Timea Pal, Abdullahi Odowa, Richard Rottenburg, Antonio Savoia, Daniel Schraad-Tischler, Laura Rodriguez Takeuchi, Géraldine Thiry, and Anna Triandafyllidou.

We would like to express our gratitude to Beth Farrow (Editorial Assistant) and Jemima Warren (Commissioning Editor in Politics and Public Policy) at Palgrave, for their support and advice in the final stages of the production process. Many thanks also go to David Frank Barnes and Francesca Iurlaro for their assistance in finalising the manuscript.

The editors and publisher thankfully acknowledge the right to re-use material that was previously published elsewhere, in particular, the OECD's permission for Chap. 9 by Guillaume Lafortune, Santiago Gonzalez, and Zsuzsanna Lonti; the European Commission's permission for Chap. 10 by Katia Berti; and the University of Chicago Press' permission to reprint an earlier article by Sally Engle Merry as Chap. 21.

Nehal Bhuta, Debora V. Malito, Gaby Umbach

Contents

Notes on Editors and Contributors

Editors

Nehal Bhuta is a Professor of Public International Law at the EUI. He is also the Co-Director of the GGP's "Global Governance by Indicators" project. He is a member of the editorial boards of the *European Journal of International Law*, *the Journal of International Criminal Justice, Constellations, and Humanity*. He edits, with Anthony Pagden and Benjamin Straumann, the Oxford University Press series in the History and Theory of International Law. He came to the EUI from the New School for Social Research, and before that, he taught at the University of Toronto Faculty of Law. Before entering academia, he worked with Human Rights Watch and the International Center for Transitional Justice.

Debora Valentina Malito is an A.W. Mellon Postdoctoral Research Fellow at the University of Cape Town. She has been a research associate at GGP (Robert Schuman Centre for Advanced Studies) of the EUI. She holds a PhD in Political Studies from the University of Milan. Her research interests focus on critical studies of liberal interventionism, state building, and governance. Her publications have appeared in International Peacekeeping, *Third World Quarterly, Peace Review*. She is working on the impact and legitimacy of the UN-NATO intervention in Libya.

Gaby Umbach is Director of "GlobalStat—Globalisation Database" and the Co-Director of the GGP's "Global Governance by Indicators" project. She holds a PhD in Political Science from the University of Cologne. Since 2015, she serves as a book review editor of *the Journal of Common Market Studies* and

teaches environmental and climate change policies and policy-making as Adjunct Professor at the University of Cologne (Germany). In 2009, she received the award of the *Stiftung Demokratie* (Cologne) for outstanding research on democratic decision-making structures within the European post- and supra-national polity. Her fields of expertise include Europeanisation and multilevel governance, employment and socio-economic policies, EU policy coordination, environmental policies, and EU enlargement and system transformation. Her current academic work focuses on sustainable development, paradigm shifts in global governance, and global governance by indicators.

Contributors

Josè Antonio Alonso is a Professor of Applied Economics at the Universidad Complutense de Madrid (Spain) and is a member of the Committee for Development Policy at the United Nations Economic and Social Council, the Development Cooperation Council at Spain, and the European Advisory Group of the Bill and Melinda Gates Foundation. His most recent books are *Global Governance and Rules for the Post-2015 Era* (with J.A. Ocampo, Bloomsbury Academic 2015); *Alternative Development Strategies for the Post 2015 Era* (with G.A. Cornia and R. Vos, Bloomsbury Academic 2013); and *Development Cooperation in Times of Crisis* (with J.A. Ocampo, Columbia University Press 2012).

Katia Berti is the Head of Sector at the European Commission (Belgium)—Directorate General for Economic and Financial Affairs—in the unit dealing with the European Economic and Monetary Union (EMU) deepening and the macroeconomy of the euro area. She holds a PhD in Economics from the University of Nottingham. She worked on fiscal policy at the European Commission, coordinating the work on fiscal sustainability indicators and public debt projections. Earlier, she worked as an economist at the European Commission's Directorate General for Employment and Social Affairs.

Robert P. Beschel Jr is a lead public sector specialist in the Governance Global Practice at the World Bank (USA). He heads the Governance and Anti-corruption in the Bank and recently released a publication called *Innovative Governance and Anti-Corruption Initiatives* that surveys some ground-breaking initiatives supported by the World Bank Group.

Christophe Blot is the Deputy Head of the Analysis and Forecasting Department at the Observatoire Français des Conjonctures Économiques (OFCE) (France).

He holds a PhD from the University of Orléans. He worked in the Central Bank of Luxembourg and then joined the OFCE, where he works on the forecasts for the Eurozone. His research focuses on the transmission channels of monetary policy, macroeconomic policies in EMU, the governance of the euro area, and the link between monetary policy and financial stability. He teaches a graduate course on European Monetary and Fiscal Policy at Sciences Po.

Paul Burger is Professor and Head of the Sustainability Research Group at the University of Basel (Switzerland). He was a lecturer at the Universities of St Gallen, Fribourg, Innsbruck, Maribor; Head of the interfaculty teaching programme People Society Environment at the University of Basel; a member of the Steering Board of Competence Center for Research in Energy, Society, and Transition and Head of its work package on Change of Individual Energy Consumption; and Head Directorate of the tri-national Upper Rhine Cluster for Sustainability Research. His research interests include theoretical foundations of sustainability (normative basis, human-nature interrelations); analysis of action spaces for societal transformations towards sustainable development; sustainable regional and urban development; energy use and supply; and methodology of sustainable research.

Alexander Cooley is Director of Columbia University's Harriman Institute for the study of Russia, Eurasia, and Eastern Europe and the Claire Tow Professor of Political Science at Barnard College of Columbia University (USA). His research examines how external actors have influenced the sovereignty, governance, and political development of the post-communist states, with a focus on Central Asia and the Caucasus. His books include *Dictators without Borders: Power and Money in Central Asia* (co-authored with John Heathershaw, Yale University Press 2017) and *Ranking the World: Grading States as a Tool of Global Governance* (co-edited with J. Snyder, Cambridge 2015).

Alyssa Dougherty is a legislative research assistant to Senator Barbara Boxer and specialises in defence, veterans, and foreign policy issues (USA). Prior to her experience on Capitol Hill, she worked as a research fellow at the World Justice Project (WJP). She completed a BA in French from Stanford University.

Hazel Feigenblatt is pursuing a PhD at the University of Maryland focusing on how the watchdog role of professional and civic media is evolving in the new information ecosystem and its impact on government accountability. In Washington, DC, she spent six years leading research projects assessing how checks and balances work (or not) in real life, such as the Global Integrity Report, the "States Integrity Investigation" (Global Integrity/Center for Public Integrity) and "Money, Politics and Transparency" (Global Integrity/

Election Integrity Project/Sunlight Foundation). She previously worked for ten years as an investigative reporter in the main newspaper in Costa Rica, where her investigations won several awards. She has previously consulted with the World Bank, the Global Initiative for Fiscal Transparency, and has held advisory roles in projects with the Mo Ibrahim Foundation, The Economist Intelligence Unit and the Transparency and Accountability Initiative, among others. She has an MSc in political science from the University of Costa Rica, an MA in public affairs reporting from the University of Maryland.

Carlos Garcimartín is a Professor of Economics at Universidad Rey Juan Carlos (Spain). He has been affiliated with the Universities of Salamanca and Complutense de Madrid. He has also worked for the Spanish Ministry of Finance and the Spanish Tax Agency. His main research topics are focused on public economics, international economics, and development economics. His articles have appeared in *The Annals of Regional Science*, *Review of Income and Wealth, Economic Modelling, Journal of Post Keynesian Economics*, *Journal of International Development,* and *CEPAL Review*.

Santiago González is the lead economist policy analyst of the OECD's publication *Government at a Glance* (G@G) (France). He received master's degrees in Economics and Public Policy. He is one of the main authors of the 2015 and 2013 editions of G@G. He is also the coordinator of the Latin American version of G@G. He worked in the Budgeting and Public Expenditures division at the OECD where he contributed to budget reviews and the development of budgeting indicators. He is a member of the Expert Advisory Group for Trust Measurement established by the OECD Statistical Directorate. He is contributing to the OECD Guidelines for Measuring Trust and to the development of governance indicators in the framework of the Sustainable Development Goals.

Daphne T. Greenwood is a Professor of Economics and the Director of the Colorado Center for Policy Studies at the University of Colorado (USA). Her articles have appeared in *Journal of Economic Issues, Review of Income and Wealth, Public Finance Quarterly, and Journal of Population Economics*. She has also been a visiting scholar at the US Treasury Department and the Institute for Research on Poverty, University of Wisconsin-Madison, as well as an elected member of the Colorado House of Representatives. Her research is focused on social costs and sustainability. Her recent books include *A Brighter Future: Improving the Standard of Living Now and for the Next Generation* (co-edited with R. P. Holt, Routledge 2014) and *Local Economic Development in*

the 21st Century: Quality of Life and Sustainability (co-edited with R. P. Holt, Routledge 2010).

Amy Gryskiewicz is a senior program associate on the research team at the WJP (USA), where she works on the *Rule of Law Index*. She holds an MA in International Relations with specialisations in Security Studies and International Law from the University of Chicago.

Maksym Ivanyna is an economist at the Joint Vienna Institute (Austria). He holds a doctor's degree in Economics from Regensburg University. He has engaged in consultancy and advisory work for the IMF, World Bank, European Commission, German Institute for Development, and Centre for EU Enlargement. His research portfolio includes publications in the fields of fiscal policy, development, growth, corruption, governance, fiscal federalism, taxation, technology adoption and innovation, and structural reform.

Guillaume Lafortune is an economist in the OECD Public Governance and Territorial Development Directorate (France). He is one of the analysts of the OECD's flagship publication G@G. He received a master's degree in Public Administration from the National School of Public Administration in Quebec. He has been involved in a number of research projects on the measurement of services to citizens and businesses, trust in government, and open government data. Before joining the OECD, he worked on a range of trade and innovation issues in the Ministry of Economic Development in the Government of Quebec.

Zsuzsanna Lonti is a senior project manager in the Public Governance and Territorial Development Directorate of the OECD (France). She is leading the work of one of the OECD's flagship publications, G@G. She holds a PhD in Industrial Relations from the University of Toronto. She is also participating in the OECD's work on Trust in Government. Prior to joining the OECD, she was a senior lecturer at Victoria University of Wellington. She also worked as a senior economist and in various policy advisor positions in the Ontario (Canada) and Hungarian governments. Her major research interests are focused on the performance of public sector organisations and human resource management in the public sector.

Sally Engle Merry is a Silver Professor of Anthropology at New York University (USA). She is also a Faculty Director of the Center for Human Rights and Global Justice at the New York University School of Law and a past president of the American Ethnological Society. Her recent books include *Gender Violence: A Cultural Perspective* (Blackwell 2009), *The Practice of*

Human Rights (co-edited with Mark Goodale; Cambridge University Press 2007), *Human Rights and Gender Violence* (University of Chicago Press 2006), and *Colonizing Hawai'i* (Princeton University Press 2000). Her most recent book, *The Seductions of Quantification: Measuring Human Rights, Gender Violence, and Sex Trafficking* (University of Chicago Press 2016), examines indicators as a technology of knowledge used for human rights monitoring and global governance. She has co-edited two books on quantification, *The Quiet Power of Indicators* (co-edited with K. Davis and B. Kingsbury, Cambridge University Press 2015) and *A World of Indicators* (co-edited with R. Rottenburg, S-J. Park, and J. Mugler, Cambridge University Press 2015).

Stephen Morse is Chair in Systems Analysis for Sustainability at the University of Surrey (UK). He holds a PhD from the University of Southampton. He is a Fellow of the Royal Geographical Society, the Royal Society of Biology, and the High Education Academy. He has a background in Applied Ecology and Environment and his research and teaching interests are broad spanning both the natural and the social sciences. These interests include methods for the assessment of sustainability in order to help guide intervention. He has also helped pioneer a number of participatory methodologies for sustainability assessment. He has been involved in research and sustainable development projects across Europe, the Mediterranean, Africa, and Asia. His book include *Rich Pictures: Encouraging Resilient Communities* (with S. Bell, T. Berg, Earthscan 2016) and *Sustainability Indicators: Measuring the Immeasurable?* (co-authored with S. Bell, Earthscan 2008).

Smoki Musaraj is an Assistant Professor of Anthropology at Ohio University (USA). She holds a PhD in Anthropology from the New School for Social Research. Her research focuses on theories of money and value, anthropology of finance, social histories of speculative bubbles, corruption, and governance, and post-socialist transformations in southeast Europe. He work has received the support from the National Science Foundation, The Social Sciences Research Council, and Ohio University Research Committee, among others. Her articles have appeared in various journals, including *Cultural Anthropology, Anthropology of East Europe Review, and Polis and Ethnologie Française*.

Moses Obinyeluaku is the chief economist and manager at the Trade and Economic Analysis Division of the International Trade Administration Commission of South Africa. He holds a PhD from the University of Cape Town. He was the Director of the Medium-Term Expenditure Planning unit in the budget office at the National Treasury of the Ministry of Finance in South Africa. He also worked as a Deputy Director of Macroeconomics in the KwaZulu-Natal Provincial Treasury and as a Lecturer in Macroeconomics and Development Economics at the University of KwaZulu-Natal. He has under-

taken research in a wide range of areas including fiscal policy, expenditure policy, economic development, and international trade. He has engaged in policy formulation and advisory work with, and for, many organisations in South Africa, including the Department of Trade and Industry, Economic Development Department, National Treasury, Cooperative Governance Department, Presidency, and the province of the KwaZulu-Natal. He also provides consultancy services to the United Nations Development Programme.

Alejandro Ponce is the Chief Research Officer of the WJP (USA). He holds a PhD in Economics from Stanford University. He joined the WJP as a senior economist and is one of the original designers and a lead author of the *WJP Rule of Law Index*. Earlier in his career, he worked as a researcher at Yale University and as an economist at the World Bank and the Mexican Banking and Securities Commission. He has conducted research and published in the areas of behavioural economics, financial inclusion, justice indicators, and the rule of law.

T.R. Raghunandan is the founder of Avantika Foundation, a not-for-profit organisation, and Centre for Decentralised Local Governance (India). He was formerly Joint Secretary—Ministry of Panchayati Raj, Government of India—and Secretary-Rural Development Karnataka before he took voluntary retirement from the Indian Administrative Service. He is now a public finance expert and an anti-corruption activist. He is a consultant and adviser to international development agencies on local government policy, fiscal decentralisation, and capacity development in local governance. With his extensive interaction with Panchayats all over the country and especially in Karnataka, he is almost synonymous with the decentralisation agenda.

Francesca Recanatini joined the World Bank (USA) in 1998 and is working on governance issues in High-Income and Fragile Countries, with a special focus on Middle Eastern and European states. She holds a PhD in Economics from the University of Maryland. She has worked on institution building, governance, and corruption since the beginning of her career working at the Center of Institutional Reforms and Informal Sector in College Park. From 2005 to 2012, she led the Anti-Corruption Thematic Group of the World Bank. She also manages the Actionable Governance Indicators work and has recently launched an initiative to promote capacity and effectiveness of Anti-Corruption Authorities around the world. She is also a member of the EU Group of Experts on Corruption.

Robert I. Rotberg is the Founding Director of the Harvard Kennedy School's Program on Intrastate Conflict (USA); president emeritus of the World Peace Foundation; Fulbright Distinguished Research Professor of International Relations at the University of Sao Paolo (2016–2017); fellow of the American Academy of Arts and Sciences; and a senior fellow at Centre of International

Governance Initiatives. His most recent books include *The Corruption Cure: How Leaders and Citizens Can Beat Back Graft* (Princeton University Press 2017), *Africa Emerges* (Polity 2013), and *On Governance* (The Center for International Governance 2015).

Anwar Shah is a Director of the Centre for Public Economics, Chengdu/Wenjiang (China), and also serves as a consultant to the World Bank and the Asian Development Bank. He has previously directed the World Bank Institute's Governance Program (2004–2009). He served in the Ministry of Finance, Government of Canada, the Government of Alberta, the US Agency for International Development and the UN Intergovernmental Panel on Climate Change. His books include *Fiscal Federalism* (with Robin Boadway, Cambridge University Press, 2009), *The Practice of Fiscal Federalism: Comparative Perspectives* (McGill-Queen's University Press 2007), and *Fiscal Incentives for Investment and Innovation* (Oxford University Press 2004).

Johannes Tonn is a Director of Programs & Partnerships at Global Integrity (USA). He leads Global Integrity's work stream on data, research, and citizen engagement, focusing on the role of adaptive learning principles. He leads a number of research projects, methodology thinking, and partnership efforts, including projects on the Africa Integrity Indicators, the State Integrity Investigation, and Global Integrity's Anti-Corruption work stream. He earned a master's degree in Political Science, Economics, and International Law at Heidelberg University. Prior to joining Global Integrity, he worked on a variety of governance projects, including social accountability-focused grassroots work with the Partnership for Transparency Fund; supporting decentralisation processes in Ecuador with the German Development Corporation, and supporting the implementation of a community-based disarmament demobilisation and reintegration project in the Niger Delta for the United Nations Office for Project Services in Nigeria.

René Urueña is an associate professor and the Director of Research at the Universidad de Los Andes School of Law (Colombia). He holds a doctoral degree in law from the University of Helsinki. He has been an associate professor at the University of Utah, a fellow at New York University, and is a docent at the Institute for Global Law and Policy at Harvard Law School.

Abbreviations

AC	Anti-corruption
AGS	Annual Growth Survey
AII	Africa Integrity Indicators
BEEPS	Business Environment and Enterprise Performance Survey
BERI	Business Environment Risk Intelligence Ratings
BTI	Bertelsmann Transformation Index
CBO	US Congressional Budget Office
CCDI	Central Commission for Discipline Inspection
CEPEJ	European Commission for the Efficiency of Justice
CF	Carbon Footprint
CPI	Corruption Perceptions Index
CPIA	Country Policy and Institutional Assessment
CSRs	Country-Specific Recommendations
DG ECFIN	European Commission Directorate General for Economic and Financial Affairs
EC	European Commission
ECA	Excess Crude Account
ECJRC	European Commission Joint Research Centre
EF	Ecological Footprint
EIP	Electoral Integrity Project
ESI	Environmental Sustainability Index
EU	European Union
FCPA	United States Foreign Corrupt Practices Act
FSI	Fiscal Sustainability Indicators
G@G	Government at a Glance
GAC	World Bank Governance and Anti-Corruption Strategy
GASB	General Accounting Standards Board
GC	United Nations Global Compact

GDA	Governance Data Alliance
GDP (PPP)	Gross Domestic Product (Purchasing Power Parity)
GFI	Global Financial Integrity
GIR	Global Integrity Report
GNI	Gross National Income
GPI	Genuine Progress Indicator
GRI	United Nations Global Reporting Initiative
HDI	Human Development Index
HPI	Human Poverty Index
HRM	Human Resource Management
IDA	International Development Association
IDRA	Institute for Development Research and Alternatives
IIAG	Ibrahim Index of African Governance
IIWB	Inclusive Index of Well-Being
ILO	International Labour Organisation
IMF	International Monetary Fund
IRCG	International Country Risk Guide
ISCO	International Standard Classification of Occupations
LAC	Latin America and the Caribbean
LICUS	Low-Income Countries Under Stress
MCC	Millennium Challenge Corporation
MNCs	Multinational Corporations
MPT	Money, Politics and Transparency
NIT	Nations in Transit
NPM	New Public Management
NSOs	National Statistical Offices
PBA	Performance Based Allocation
PEFA	Public Expenditure and Financial Accountability
PFM	Public Financial Management
PGC	Public Governance Committee
PISA	Programme for International Student Assessment
QRSs	Qualified Respondent Questionnaires
SD	Sustainable Development
SDGs	Sustainable Development Goals
SFI	State Fragility Index
SGI	Sustainable Governance Indicators
SI	Sustainability Indicators
SII	State Integrity Investigation
SNA	System of National Accounts
SWDs	Staff Working Documents
TI	Transparency International
UN	United Nations
UNCAC	United Nations Convention Against Corruption Assessments

UNDP	United Nations Development Programme
US	United States
WB	World Bank
WCED	World Commission of Environment and Development
WEF	World Economic Forum
WGI	Worldwide Governance Indicators
WJP	World Justice Project
WVS	World Values Survey

List of Figures

List of Tables

List of Boxes

1

Introduction: Of Numbers and Narratives—Indicators in Global Governance and the Rise of a Reflexive Indicator Culture

Nehal Bhuta, Debora Valentina Malito, and Gaby Umbach

Knowing, Governing, and Changing the World with Quantification

Since the late seventeenth century and the rise of the early modern state, *knowing* and *governing* have been intimately connected. An important justification for the extraordinary concentration of power and prerogative that marked the emerging absolutist state was the idea that such power was instrumental to the *reasons* of state (*ratio status* or raison d'état)—the peace, prosperity, and well-being of the population of the state. Realising reasons of state required knowledge of a population and a territory, in order to husband its productive resources, perpetuate and develop its moral order, and raise the revenues of the state (Tribe 1995, pp. 1–8). In order to be better regulated, the social body—understood perhaps for the first time as an aggregate of persons or a *population*—had to be *known*. Counted, certainly, but also described, classified, compared, and so made legible to administration and governing. Statistics was born as *notitia rerum publicarum* or

N. Bhuta (✉) • G. Umbach
European University Institute, Florence, Italy

D.V. Malito
University of Cape Town, Cape Town, South Africa

D.V. Malito et al. (eds.), *The Palgrave Handbook of Indicators in Global Governance*,
https://doi.org/10.1007/978-3-319-62707-6_1

Staatenkunde, later *Statistik* (Lindenfeld 1997, pp. 11–45; Foucault 2007, pp. 100–101, 314–315), a component of the sciences of the state, the purpose of which was the rationalisation of governance (Desrosières 1998, pp. 16–44). The rationalisation of knowledge *of* the state also had a role in preserving peace *between* states in Europe: the effective preservation of the balance of power required that "each state is in a position … to know its own forces, and … to know and evaluate the forces of others, thus permitting a comparison that makes it possible to uphold and maintain the equilibrium" (Foucault 2007, p. 315).

Governing the state with and through knowledge of the state—that is, with statistics—was an essential part of *the making* of the modern state (Bowker and Leigh Star 1999, p. 118). Efforts to create consistent and coherent knowledge of a population and a territory also helped create the administrative, legal, and economic uniformity that came to be identified with *what it meant to be* a modern and rationally governed nation-state:

> The universality and clarity of the system of weights and measures enabled people to avoid deceit in commerce and trade, whereas the administrative and judicial forms of encoding were indispensable in lending an objective consistency in things that could not otherwise by counted: marriages, crimes, suicides, and later on, businesses, work-related accidents, and the number of people unemployed. …
>
> … The most visible manifestation of this process of homogenization and codifying many aspects of human existence was the unification of the national territory, since many things and rules that were then redefined and generalized had previously been specified at a local or provincial level. (Desrosières 1998, p. 32)

In the peace after Napoleon, European states greatly expanded their bureaucracies and institutions (Desrosières 1998), and with the liberal demand for more rational, enlightened, and productive government of people and a territory (Poovey 1998, p. 147)[1] came the demand for more statistical knowledge collected and (over the course of the nineteenth century) made available for debate in the public sphere (Hacking 1990, pp. 16–26, 35–46). The collection and dissemination of statistical information went hand in hand with reformist ambitions for rendering government more effective in its interventions *on society* in relation to problems of "moral degeneracy" such as crime, suicide, and prostitution. Deviancy and normalcy could assume seemingly objective and scientific meaning only with repeated attempts over the century to quantify (and define) concepts such as mortality, sickness, health, and

crime (Hacking 1990, pp. 47–54, 95–104). The result was "an avalanche of printed numbers" (Hacking 1990, p. 138):

> Nothing was left untouched by the statistician … During the years 1820–1840 the rate of increase in the printing of numbers appears to be exponential whereas the rate of increase in the printing of words was merely linear … [T]he avalanche is not merely a quantitative fact but a change in our feeling about the sort of world we live in.
>
> … Equally important was the subtler, subversive influence of the new group of human categories coming from this avalanche of numbers. (Hacking 1982, pp. 281–2)

The *categorising* and *counting* essential to creating this form of knowledge brought into existence new objects. *Quantification* through categorising and counting converts into numerical existence what was previously expressed in descriptive words (Desrosières 2015, p. 333). *Measurement* implies that a quantum or object exists which can be measured. A social reality that is not already quantified through the prior hermeneutical labour of categorising, defining, and counting is not readily measurable. Thus, as historians of statistics have repeatedly pointed out, the *quantification* of concepts such as "unemployment", or even "cause of death", (Bowker and Leigh Star 1999, pp. 53–106) entails extensive and intensive processes of negotiating and fixing meanings and creating equivalences (Espeland and Stevens 1998, 2008) of what may be quite diverse events and phenomena:

> The verb 'to quantify,' in its active form (to make into a number), supposes that a series of hypothetical equivalences has been developed and made explicit, involving comparisons, negotiations, compromises, translations, registrations, encoding, codified and replicable procedures, and calculations leading to numericization. *Measurement itself comes after that*, as the rule based implementation of *these conventions*.
>
> […] Statistics and all forms of quantification in general … change the world through their very existence, their circulation and their rhetorical usage in science, politics or journalism. Once quantification procedures are encoded and become routine, their products are objectified. They tend to become reality in an apparently irreversible way. The initial conventions are forgotten, the quantified object is naturalized and the use of the verb 'to measure' comes to mind or is written with no further thought. (Desrosières 2015, pp. 333–4)

The quantification and measurement of the social, political, and economic is thus a means of generating new ontologies of the real. The objects produced

through quantification and measurement are not exactly fictions, but neither can they be understood through a naive "metrological" realism as straightforward representations of an underlying reality. As observed by Desrosières (2001, p. 348), the seductive but deceptive character of quantification and statistical knowledge is their invitation to the conclusion that "the computed moments (averages, variances, correlations) have a substance that *reflects an underlying macrosocial reality*, *revealed by those computations*. Therein lies the essence of metrological realism".

But it is this metrological attitude towards the social, which quantification and statistical reasoning helped produce and perpetuate, that makes statistical knowledge at once instrumental to, and productive of new problematics for, governing. Within the historical context of state-making and the intensification of state power, statistics were a vector of acting by knowing: classifying, counting, aggregating, and averaging all produced a common language that allowed disparate phenomena to appear as solid objects available to social actors or researchers, because the objects are embedded in "descriptive and explanatory systems … capable of orienting and triggering action" (Desrosières 1998, p. 248). Quantification and statistical knowledge are a means of governing and coordinating collective action at the level of the nation-state (in the first instance, at least), because collective action across large spatial and human scales requires correspondingly extensive and seemingly de-localised kinds of knowledge. This *kind* of knowing seems to allow players to place macro-realities (the national rate of unemployment, the rate of infant mortality, the total production of an entire national economy) on stage *and act upon them* (Desrosières 1998, pp. 199–209; Speich 2011).

At the same time, the style of reasoning of statistical science and its underlying metrological claims (Hacking 2002, pp. 178–199) are especially susceptible to reflexive critique both from *within* (by statistical experts challenging the internal validity and reliability of the claims according to available logics of statistical inference) and *without* (by specialists of different kinds of knowledge production such as journalists, ethnographers, historians, lawyers, philosophers, and many others: Hacking 1990; Porter 1995, pp. 89–113, 148–190). The result is that while truth claims and claims of objectivity made through quantification and statistical knowledge are often presented as circumscribing their own contestation or hiding value judgements behind technical operations and specialised methods, this moment of uncritical acceptance is short-lived, if it occurs at all. As soon as the interests, authority, and influence of some set of social actors are implicated in the metrological claim of the statistical measure, contestation is likely to emerge. In this sense, as Porter

observes, quantification and its associated processes of standardisation were understood in the nineteenth century in France, Britain, and the United States as opening governmental decision-making to wider scrutiny by giving it foundations in apparently objective and calculable knowledge claims, rather than the closely held and inestimable "judgement" of a certain kind of official (an aristocrat or established family scion). Objectivity was supposed to replace unaccountable discretion, and the basis for objective decision-making could be established through scientific methods:

> [T]his faith in numbers was wedded to a belief in progress through public information. A science of statistics based on subtle arguments and requiring long experience was poorly calculated to influence public debate, or to provide a justification for public decisions … Statistics was supposed to provide thoroughly public knowledge, suitable … for a democracy. Ideally, democratic statistics would be self-explanatory. (Porter 1995, p. 80)

But the esoteric *savoir-faire* of the social superior would be replaced with the technocratic judgement of the expert. Quantification and statistical reasoning would not, if transpired, enable the transcendence of politics and of discretion but its *displacement and transformation* into another sphere of specialised knowledge. Technocratic ideals may have promised a mechanical objectivity of decision-making, but rarely realised it, not the least because those working within professional communities of expertise of quantification had a practical sense of the many uncertainties, decision points, normative and empirical assumptions, and theoretical arguments hidden away within the processes of making a discrete number that represents a complex social reality (Speich 2011).[2] Measurement rests on convention, and conventions are challengeable and frequently provisional as "best efforts" to conceptualise and count something, which is not inherently countable. But the knowledge needed to expose and render transparent the conventions, judgement calls, devices of statistical reasoning, and sources of information that go into producing a quantified object is itself highly specialised and largely within the domain of disciplinary scientific communities (Espeland and Stevens 2008; Desrosières 1998, 2015).

For those outside such communities, or otherwise not well trained in piercing the black boxes upon which processes of quantification and measurement rely, a certain "trust in numbers"—or at least, one might say a trust in the idea that the number *bears some relationship* to the object it purports to measure—seems to become a strong inertial tendency of contemporary social, political, and economic discussion.

Global Governance and the Salience of the Contemporary Avalanche of Indicators

There is a pervasive sense in which we seem to be living under a new avalanche of numbers, and in particular an avalanche of indicators *beyond the state* and purporting to create knowledge on a *global scale*. Anthropologist Sally Merry (2016, p. 3) observes that the production of, and interest in, global indicators is "now booming". Political scientist Alexander Cooley refers to a "frenzy" of ratings and rankings and identifies 95 rankings, ratings, and indexes—the overwhelming majority of which having emerged after the end of the Cold War and as economic globalisation accelerated (Cooley 2015; The World Commission on the Social Dimension of Globalization 2004; Sassen 2007, pp. 11–96). International relations scholars Judith Kelley and Beth Simmons (2014) record over 160 "global performance indicators", with more than 8 new ratings and rankings added on average per year since 1999. Just as quantification and measurement went hand in glove with the rise of the modern state, indicators have become a highly salient feature of our post-national architecture of governance.

In ordinary language, an indicator is simply a thing or word that points to something else. A "numerical indicator" in its early-twentieth-century usage was an output of a process of classifying and counting that *pointed to*, *signalled*, or in some sense *showed* an underlying phenomenon in economy and society: business activity, agricultural productivity, the mental health of the population, and so on (Porter 2015). To say that the recorded changes in the level of corn prices, hog prices, or blast furnace output were indicators of the level of business activity was to maintain that a specific outcome of classifying and counting amounted to a representation of a more complex general reality that itself cannot be counted. An indicator is a number that *stands in for* that general reality and purports to represent it; it is not only an expression of what is actually classified and counted. *Standing in for* is not the same as measuring: in the example just given, corn prices are logged, averaged, and measured. Business activity is not. However, the metrological attitude described above, which is closely intertwined with the rise of modern statistics and quantification, can blur the line between our understanding of "standing in for" and "measuring". Through the development of the indicator, the underlying messy social world is given a concreteness and discreteness that seems to make it once more knowable and theorisable, as well as more easily acted on, intervened in, and governed. It is perhaps because of these powerful "knowledge-effects" and "governance-effects"

(Merry 2016, pp. 4–5) that over the last decade the production, use, and misuse of indicators at a global scale have attracted so much attention and debate.

A crucial feature of the contemporary explosion of indicators and rankings is their extraordinary proliferation of subject matters, institutional authors, and methodologies. From economic openness to human development, gender empowerment, democracy, corruption, budget transparency, rule of law, state fragility, and innumerable other concepts and policy buzzwords, almost anything appears to be amenable to quantification, aggregation, and ranking (Cooley and Snyder 2015, pp. 194–203; Kelley and Simmons 2014). Many such indicators, like the United Nation's Development Programme's Human Development Index (HDI) or Freedom House's Freedom Index, aim not simply to quantify and measure an existing social reality but to enhance the recognition and salience of a concept or policy objective as an alternative to others (Bradley 2015). In the well-documented case of the HDI, the ambition was to create a simple and attention-grabbing measure that would highlight the importance of other dimensions of human welfare not addressed through a focus on aggregate income growth (Fukuda Parr 2003). As such, indicator production has an implicit (or, in some cases, explicit) normative ambition: quantifying a concept (corruption, sustainability, governance, gender empowerment, etc.) is often about making that concept *more salient* and *more visible*. To the extent that quantification lends tangibility to something otherwise hard to define or agree upon (e.g. human development), the number or measure comes to *exemplify* the concept and instantiate its realness (Desrosières 2001, p. 348).[3] Hence, an important part of what is at stake in the new avalanche of numbers is not just whether or not an indicator is a good or bad number from the perspective of measurement theory and statistical science. Equally important is whether the concept it instantiates becomes part of how we identify, frame, construct, and respond to global problems within discourses of global policy and donor action: a policy platform that takes as its understanding of development the promotion of "GDP growth" may look very different and have very different consequences to one which defines development in light of the promotion of "human development".

Importantly, one of the essential ways in which indicator producers aim to achieve greater salience for an idea, concept, or normative objective is by purporting to *rank* public and/or private actors' achievement of this objective. Quantification and measurement in this "world of indicators" (Rottenburg et al. 2015) is closely connected to the evaluation and auditing of behaviour. For this reason, contemporary global indicators are sometimes described as "benchmarks" or "performance indicators" (Broome and Quirk 2015a, b),

inasmuch as they purport to monitor and rate states' "relative performance … in a given policy area" (Kelley and Simmons 2014, p. 4).[4] Kelley and Simmons define "global performance indicators" as "public, comparative and cross-national indicators that governmental, intergovernmental and/or private actors use regularly to attract attention to the relative performance of countries in a given policy area" (Kelley and Simmons 2014, p. 4). Davis et al. (2012, pp. 73–4) develop a working definition of "indicator" that also emphasises the importance of performance measurement, comparison, and ranking in light of some standard:

> An indicator is a named collection of rank-ordered data that purports to represent the past or projected performance of different units. The data are generated through a process that simplifies raw data about a complex social phenomenon. The data, in this simplified and processed form, are capable of being used to compare particular units of analysis (such as countries, institutions, or corporations), synchronically over time, and to evaluate their performance by reference to one or more standards.

Cooley (2015, p. 13) differentiates between ratings—which assign a discrete value or grade to the performance of a state—and rankings, which are inherently relational "as states are assigned an ordinal ranking in comparison to each other".

Far from exercises in academic knowledge production, indicators are thus widely perceived to be closely connected with the *exercise of power* within and between states, and thus with the global governance of public and private actors. In this context, recent contributions speak of *Scorecard Diplomacy* (Kelley 2017), *The Politics of Numbers*, and *How Numbers Rule the World* (Fioramonti 2014), while others refer to rankings as a "Tool of Global Governance" (Cooley and Snyder 2015) or evoke the *Quiet Power of Indicators* (Merry et al. 2015). *How* one grasps the nature of this power depends heavily on the disciplinary framework of analysis being used. Anthropologists and sociologists conceive of the contemporary power of numbers at the global level in terms of the "linkage between knowledge and power" (Merry et al. 2015, p. 1), where indicators are "part of a regime of power based on the collection and analysis of data and their representation … Rather than revealing the truth, indicators create it", shaped by inequalities of power and expertise (Merry 2016, p. 5). Scholars working in this vein tend to trace the genealogies of specific indicators in order to unravel the underlying normative orientations and ambitions of the processes of quantification; the kinds of institutional or other authorities that underwrote the legitimacy of the indicator's

claim to objectivity; and the uncertainties, ambiguities, objections, and controversies that had to be overcome or somehow left behind in order for the indicator to successfully come to be seen to represent a reality that was being measured (Merry and Wood 2015).

Scholars working in these methods also understand the effects of such knowledge production in a broad way, tracing the impact of indicators by examining who uses them and how and in what contexts they achieve or fail to achieve uptake and influence. Often influenced by work in social studies of science and technology (Merry et al. 2015, p. 5), such richly contextual and intensively qualitative empirical studies of the power of indicators resist easy generalisation. Simple hypotheses about the power of numbers or the impact of quantification on public and private action are not straightforwardly validated or disproved. For example, while Bukovansky (2015) emphasises the influence of global corruption rankings such as Transparency International's Corruption Perception Index (CPI), Migai Akech's (2015) detailed study of the CPI in Kenya shows that it was perceived as inaccurately and inappropriately focusing on petty corruption rather than grand corruption, diminishing its credibility among anti-corruption constituencies. But the findings of such scholarship do show that any understanding of the power of numbers and indicators in contemporary global governance requires attention to and study of the *processes* and *pathways* of knowledge creation and circulation that connect territorially embedded "global" centres of knowledge production and calculation (such as international organisations in New York, Geneva, or Brussels), with national, regional, or local sites of policy-making, politics, and power. This disciplinary perspective on indicators is concerned to better understand "the relationship between global systems of knowledge production … and the local meanings, uses and effects of these global systems" (Merry et al. 2015, p. 17). Under some conditions and in some contexts, these effects can reproduce inequalities in wealth, influence, and expertise, between and within developed and developing countries. As Merry observes:

> Insiders with skills and experience have a greater say in developing measurement systems than those without - a pattern that excludes the inexperienced and powerless. At the global level, experts are usually cosmopolitan elites with advanced education or people who have had previous experience in developing indicators of the same kind. They are often from the global North and trained in political science, economics, or statistics … Countries that have carried out the relevant surveys create the models for the next set of surveys … In the context of global governance, this means when experts gather to develop indicators and plan data collection, those from countries that have already tried such data gathering and

> analysis projects claim special knowledge and authority. … Local vernacular knowledge is typically less influential than more global, technical knowledge… Classification systems often grow out of local systems of knowledge that become globalized systems into which the local systems of other countries must be squeezed. (Merry 2016, p. 5, 21)

Political scientists and international relations scholars label the foregoing approaches as "Foucauldian" and argue that while they are relevant, they do not capture with sufficient generalisability the range of ways in which indicators influence behaviour at the global level. Cooley proposes a functional categorisation, claiming that indicators, ratings, and rankings function in four distinct ways as a means of governance at the global level (Cooley 2015, pp. 14–23).

First, they act as *judges* by classifying and labelling the performance of states and other entities relative to each other (in the case of an index) or on a scale (in the case of a rating). Judging could elicit rejection of the judgement and contestation of the indicator by the entity being judged, leading often to a critique of the indicator's conceptual and empirical validity or a rejection of the very standard against which the indicator purports to evaluate the entity. But, depending on the power and influence of the entity being evaluated, and the authority and legitimacy of the indicator producer and the indicator itself (Kelley 2017; Dutta 2015),[5] the judgement could have direct or indirect effects on the state's national policies and conduct. In the case of the World Bank's Ease of Doing Business Indicator, several states have actively sought to improve their ranking by adopting targeted reforms that are calculated to raise their overall score (Schueth 2015), while public discussion of a negative ranking could also result in strategic policy responses in order to be seen as "doing something" about the judgement.

Second, indicators and indexes can become integral to *global regulatory practices*, both formally and informally. They may be included as reference points for assessments of compliance with mandated standards (Dutta 2015; Serban 2015), or as hard benchmarks to be met for funding eligibility. They can also exercise regulatory functions in more subtle ways: by transforming local or particular models of governance and normative frameworks into apparently global technical standards and benchmarks (good governance, the rule of law, sustainability), indicators, rankings, and ratings can become technologies of policy transfer and diffusion (Bulmer and Padgett 2004, p. 5; Knill 2005). Laws and policies that conform to or approximate a given historical experience or prototype (often those of the contexts of origin of the indicator and its underlying concept) score well on the index because they

evidently demonstrate a strong fit with what the index aims to measure (Bhuta 2012; Broome and Quirk 2015b).

Third, indexes and rankings can be developed and used expressly as *advocacy tools* by issue activists and networks, in order to make visible, highlight, and frame a problem such as corruption. The rankings and ratings here act as part of a naming and shaming strategy in respect of the state or other entity, drawing attention to poor (or, commending successful) performance on a specific issue relative to rivals. Indeed, for some indicator producers, the point of quantifying and ranking is to promote change in policies and awareness on the issue, and to have the number become a focal point for social pressure on the leadership of the state (Kelley and Simmons 2014). The fourth role played by indexes and rankings is closely connected to the third: Cooley notes that producing a quantified measure of performance in relation to a particular issue or policy norm is a way in which the organisation creating the *measure "flag plants" or "brands"* itself as a relevant participant in debates over the issue. Perhaps one of the most successful instances of such "flag planting" is Transparency International's Corruption Perception Index, which became closely associated with the authority and expertise of the organisation itself.

The origins of the contemporary avalanche of numbers lie in shifting realities and theories of government and governance. It is a commonplace to remark that we live in a period of "global governance", but the features of this contemporary reality (Hurrell 2007, pp. 95–120) can be concisely summarised as follows: a deepening of inter-state governance which relies on formal and informal processes and institutions to manage common and interdependent problems through cooperation and coordination, rather than command and control (the latter being associated with the regulatory mode of government of the sovereign state within its own jurisdiction). This deepening of inter-state governance, in some cases within "multilevel, multilayered and multitiered political system[s]" (Umbach 2009, pp. 45–46), does not only employ classical power political methods of bargaining and consent-based obligation, it also increasingly relies on modes of "complex governance" that create and operationalise functional and expertise-driven norms, practices, and communities of knowledge that link public and private actors, "political" and "expert" authorities, as well as legal and non-legal modes of normativity through networks and hybrid bodies. Resulting from an increasing globalisation and regionalisation of interactions, these phenomena materialise the integration and "extension of political structures and arenas beyond [and across] the nation state" and international organisations as manifestations of modern complex governance (Umbach 2017, p. 47). This complex governance beyond the state comprises *both* recognisably public legal-normative

authority, such as that exercised by international courts, human rights treaty bodies, as well as the executive organs of international organisations (such as the Security Council or the Council of Europe), *and* "private authority structures …, private systems of arbitration and dispute settlement, privatized rule production resulting from technical standardization" and public-private regimes governing particular sectors of the global economy (Hurrell 2007, p. 109).

A structural feature of the shift from government to governance—that is, from regulatory command and control characteristics of ruling within sovereign states to cooperation and coordination between and across sovereign states and to incorporating other public and private actors—is that interactions between actors exercising diverse kinds of authority and power must be managed through "the production, management and regulation of knowledge and information" (Voß and Freeman 2016, p. 4). Collective action, collective order, and collective modes of seeing, judging, and acting cannot be created by fiat or decisive coercion at the global level (in the absence of a true global empire). Moreover, just as government at the scale of the nation-state requires the production of knowledge that facilitates knowing and acting at a *national* territorial scale, governance at the global level demands forms of knowledge that are sufficiently stripped of national and local contexts to facilitate comparison, judgement, and action across national and regional scales. Quantification in the form of indicators facilitates the creation of such global-scale knowledge. Numbers invite comparison, and comparability of contextual social and political phenomena such as "corruption" requires a considerable labour of standardisation and commensuration (Espeland and Stevens 2008)—one which of necessity requires ways of talking about a phenomenon as if it were in some sense the same in different places and despite its locatedness and specific history in place A, as distinct from place B (Merry 2016, p. 212).[6] Global policy and global action require "distributed cognition", in which "a large number of parties coordinate their work to maintain some larger order or enterprise" (Merry 2016, p. 418). Thus, Voß and Freeman (2016, p. 5) observe that:

> It is in transnational governance beyond the state that shared ontologies, rationalities, models and technical standards of governing often develop momentum as an independent force of collective ordering …
>
> … It is in the transnational sphere that the use of monitoring mechanisms, benchmarks, guidelines and new mechanism of participation is especially notable.

In the modes of developing a (often fragile) collective order that marks global governance, the making of knowledge *about* governance is a medium *of* governance: "established representations of the reality of governance form the basis on which politics … operates … Knowing about governance, like this, is at the same time also governance by knowing. It is a way of shaping political conduct" (Voß and Freeman 2016, p. 22). Relatedly, Merry (2016, p. 10) points out that one of the other driving forces behind the increasing importance of quantification in global governance is the emergence of so-called new (modes of) governance. Such new modes of governance represent "a broad range of regulatory strategies that rely on empiricism, quantitative knowledge as the basis for decision-making, and problem solving through benchmarking. Key features are […] stakeholder participation, flexibility, reversibility, monitoring and peer review, transparency […] and learning oriented and multilevel decision making" (de Burca 2010; de Burca et al. 2014).

Another important, and perhaps more prosaic, source of the rise of indicators is a transformation of government *within* the state under the pressure of what can broadly be labelled "neo-liberal" (Collier 2012) practices and theories of economics and of government.[7] The post-Keynesian, post-1970s demand for the rationalisation of public expenditure and government regulation was paralleled by the rise of "new public management" (NPM) and "performance-based management" techniques that aimed to induce processes of regular self-evaluation and self-improvement within workers and managers across public and private bureaucracies (Power 1997; Roberts 2010; Strathern 2000). Audit cultures aim to responsibilise the evaluated, incentivising them to adopt calculative practices that aim at "doing better on the measures" rather than working to a hierarchically imposed rule or principle. Power notes that NPM "emphasizes cost control, financial transparency, the autonomization of organizational sub-units, the decentralization of management authority, the creation of market and quasi-market mechanisms … and the enhancement of accountability to customers … via the creation of performance indicators" (Power 1997, p. 43). These characteristics of public spending reform generated demand for "an intensification of financial and non-financial information flows. The 'hollowing out of the state' by the NPM generates a demand for audit and other forms of evaluation and inspection … [and requires] specific technologies of reaggregation and recentralization which accounting and auditing promise" (Power 1997, p. 44; Desrosières 2015, p. 334, 337, 338).

Indicator Cultures and the Handbook of Indicators in Global Governance

In her recent monograph, anthropologist Sally Merry has coined the term "indicator culture" to capture the contours of the recent explosion of indicators, rankings, and ratings (Merry 2016, p. 220). For Merry (2016, p. 221), an indicator culture describes:

> A set of techniques and practices of knowledge that has acquired a significant level of public trust and acceptance. Both in policy circles and in the general public, there is a faith that numbers and scores can provide secure knowledge of a world that seems unknowable … In this era of quantitative enthusiasm, we use data to define problems and construct solutions to them … Indicators seductively promise to provide guidance through a complex world.

The idea of a "culture" of indicators is helpful as it captures not only the enthusiasm for and expanding production of indicators but also the idea that this new avalanche of numbers has brought with it particular dynamics of knowledge production, counter-knowledge production, validation, contestation, trust, and scepticism. In light of the deeper changes in governance discussed above, our contemporary indicator culture is unlikely to disappear any time soon. Even as complex governance and new public management endeavour to replace "political" command and control with "technical" standards and expert-driven "problem-solving" methods, we also seem to be living in a period in which numbers' objectivity and validity are quickly subjected to criticism and contestation. Espeland notes that one of the interesting features of the pervasive production and use of indicators these days to make a claim about how the world is or how it works is that those affected or evaluated by the claim *react* with counter-claims that can take the form of alternative numbers, or new "stories about what [the indicators] mean, what … their virtues or limitations [are], who should use them to what effect, their promises and their failings" (Espeland 2015, p. 65).

Thus, it is important to recognise that as much as our indicator culture engenders a "faith in numbers", the very expansion of the power of numbers and their role in (global) governance over the last 20 years has brought with it a heightened sense that quantification, indicators, and rankings are a way of doing politics that must be engaged with from within and without the specific disciplinary knowledges (such as statistics and econometrics) that underwrite their claims to objectivity. The massive expansion of critical and reflective

scholarly writings *about* quantification and indicators over the last 10 years (much of it cited in the references to this chapter and other chapters in this Handbook) is perhaps the clearest testament to a strong trend towards critical engagement with the uses of numbers and measures, and it is reasonable to think that this growing scholarly interest taps in to a wider dynamic of *reflexivity* that also forms part of our contemporary indicator culture. This is not to suggest the many claims made for the power of numbers and the impact of quantification are unfounded; rather, it is to propose that any book, which approaches the "state of the art" of indicators in global governance, must attempt to capture this emerging cultural reality that encompasses *both* a constant turn to quantification as necessary—even inevitable—for policy-making and decision-making in global governance *and* the fractious and contested nature of the authority of any given outcome of quantification.

The chapters collected in this Handbook comprise such an attempt to capture the contemporary indicator culture, with all its discordant and contrasting orientations. The chapters are contributed by those who are active participants in the world of indicators—*producers*, such as international organisations and non-governmental organisations, and *consumers* as well as *meta-consumers* such as scholars and policy consultants. The Handbook focuses on three broad kinds of indicators in global governance, for the relatively simple reason that these are among the most debated and discussed ones: indicators relating to corruption, indicators relating to governance and the state, and indicators relating to sustainability.

Before moving to short accounts of the chapters themselves, some of the themes can be helpfully identified.

Post-Metrological Realism and Reform

No chapter of this Handbook adopts a naively metrological understanding of indicators as simply "measuring" reality. All authors could rather be called "post-metrological realists" in greater or lesser degrees. From the perspective of contributing indicator producers, the idea that the indicators measure *something* is indispensably connected to the view that indicators are valuable because they may be *actionable*, generating comparative or locally relevant information which is sufficiently reliable to help identify problems and understand possible solutions. As in the past, the promise of quantification is a promise of diagnosis that can guide reform; it potentially allows the nature of a problem to be better understood. An interesting theme running through the

contributions by producers is a consistent criticism of an earlier generation of indicators that purported to be aggregate measures, composites of other data sources that could readily be turned into a ranking.

An emerging trend that is clear in these chapters is a criticism of such aggregate measures, *because they are not actionable* and do not provide any "policy-relevant" information that could be used to target reforms. Instead, there is a turn to developing disaggregatable, local experience-based measures that may not be comparable across countries, but could guide reform measures. Similarly, "dashboards" rather than "league tables" are advanced as a way of generating a multidimensional picture of the phenomena. One can see a clear responsiveness to many of the criticisms of a first generation of global-scale indicators and metrics in these chapters, along with a defence of the utility of quantitative data as a source of insight and actionable knowledge. An awareness of the political sensitivity of rankings is also evident, and an undertone in the chapters from indicator producers is that rankings and indexes are not only not very useful for formulating concrete reform measures but may also be counterproductive because of the political blowback they can generate.

Limitations of Concepts and Data as Inherent and Subject to Contestation

The normativity of measurements, and the inevitability of importing a theory or model of the phenomena that sought to be measured, is a consistent theme of the contributions by scholars and critics of indicators—and is indeed frequently acknowledged even by producers. Concepts like corruption, rule of law, stateness, and fiscal or environmental sustainability have no natural or non-theoretical meaning. To attempt to measure them is to develop ascriptive and prescriptive criteria for their realisation or non-realisation, and also to develop a set of assumptions about what constitutes factual or observable evidence for their presence or absence. Similarly, the data that is used to quantify the phenomena will more often than not be attempts to count what can be counted, in the form of proxies for the phenomena. The relationship between the proxy and the wider phenomenon itself must ultimately rest on theoretical foundations that may also be strongly normative. One of the perils of quantification of corruption, governance, and sustainability consistently discussed by scholars in this volume is hence that of the normalisation and reification of very particular normative commitments and historical experiences, to the exclusion of alternative theories, norms, and historical possibilities. The result is that indicators which "black box" these commitments and

experiences beneath seemingly technical and factual claims about their construction may help advance or be complicit in specific political and economic agendas.

This set of criticisms can generate several different responses, as seen from contributions in Part II of this Handbook. One is to demand alternative measures (Chaps. 13, 14, 18) that bring in alternative theories of the phenomena and content, which can be observed most clearly in the chapters concerning fiscal sustainability, but also in debates about measuring corruption and governance. Another response, which to some degree is reflected in the contributions by indicator producers, is to eschew highly general conceptions of the phenomena and look instead for defining the object of quantification in more context-specific ways and using more "granular" data (Chaps. 8, 9, and 10).

The price of this richer contextualisation of a concept is that the demand for simple, aggregated, global level comparisons of a phenomenon cannot be met. However, one of the emerging qualities of our indicator culture may be that among policy-makers and professional consumers of such information, greater awareness of the limits and utility of aggregate and composite indicators is generating a willingness to "drill down" to lower levels of generality and accept that context-specific knowledge is essential in grasping the relationship between an indicator and the underlying phenomena that it "stands in for". This in turn suggests that the "data publics" presupposed by a new generation of indicators must be tolerant of greater complexity in deriving conclusions from indicators, and be willing to accept and make explicit the extent of interpretation and judgement entailed in using such measures. In other words, a turn to more contextually constructed measures of corruption, governance, and sustainability could lead to a greater role for complementary forms of knowledge such as the qualitative and ethnographic—and to a necessary acknowledgement of the *discursiveness* and *narrativity* involved in using quantitative measures to draw a picture of the world.

Complex Effects of Indicators

A third theme that cuts across the chapters of the Handbook is that the relationship between indicators and governance is complex. Indicators are used by a variety of actors for a variety of purposes. Despite their popularity and seeming capacity to shape debate and perception of a problem or issue, the power of indicators remains highly contextual and dependent on how they are enrolled in particular, situated networks of actors and influence. Quantification is a potential medium of power, but more factors must be present before we

can say with certainty that the numbers or the process of producing numbers have had an impact. A converse implication is that easy assumptions about the utility of quantification in shaping a reform agenda or changing a public debate may be falsified.

Chapters in This Handbook

This Handbook addresses the theoretical and analytical issues raised by the production and consumption of indicators across three areas most relevant for global governance activities: corruption and anti-corruption policy, fiscal sustainability and sustainable development, and governance and stateness. It is a collection of chapters written by both theorists and practitioners in the field of indicators. The chapters by Beschel (Chap. 7); Recanatini (Chap. 8); Lafortune, Gonzalez, and Lonti (Chap. 9); Berti (Chap. 10); Dougherty, Gryskiewicz, and Ponce (Chap. 11); as well as Feigenblatt and Tonn (Chap. 12) are contributions from current or former producers of indicators.

In Chap. 7, Robert P. Beschel reflects on the many limitations of existing comparative indices of governance, corruption, and transparency, looking back on 20 years of debate concerning the measurement of governance. He notes the widespread criticisms of common governance measures for their lack of concept validity and heavy reliance on subjective data sources. Beschel argues that some of these concerns, and the additional criticism of aggregate governance measures as unsupportive of developing concrete reform measures, were addressed in a second generation of measures that focused on more granular comparative assessments of dimensions of government performance such as public expenditure system evaluations and anti-corruption treaty compliance reviews. Notwithstanding the criticisms of indicators, there is a continued interest in refining and developing them to help assess progress towards global policy goals such as the Sustainable Development Goals. Beschel (p. 15) contends that even as rankings of corruption and transparency were to be viewed with some scepticism, developing systematic sector evaluations (potentially not comparable across countries) can nonetheless be a means to "generate a more detailed menu of actionable reforms that can help guide policy makers".

In a somewhat similar vein to Beschel, Francesca Recanatini observes in Chap. 8 that the first generation of corruption measures failed to grasp the heterogeneity and multidimensionality of governance and corruption and did not provide guidance on where best to address reforms or how to sustain them. She reflects on the theoretical insights that led to a demand for different

ways of measuring these phenomena, using experience-based surveys disaggregated by regions and cities and conducted through a participatory method. The results, as she notes, cannot be compared with other countries, but can generate very specific reform recommendations which address specific organisational features of governance within a given country.

In Chap. 9, Guillaume Lafortune, Santiago Gonzalez, and Zsuzsanna Lonti criticise composite indicators as potentially misleading and prone to simplistic analytical or policy conclusions. They argue for a "dashboard" approach which—in light of the lack of consensus on a general concept of good governance—presents different sources and data next to each other. Instead of aggregating and weighting data into one single index, they elaborate on the example of the OECD's "Government at a Glance" approach, which aims at building valid and reliable indicators that are actionable and robust and which illuminate specific dimensions of public administration within a state. The authors however acknowledge that the dashboard approach is not "media-friendly" and that it may thus not be as widely known in the public. They also argue that the indicators developed in this way provide greater insights into the direction and required content of reforms to policy-makers.

Katia Berti discusses in Chap. 10 the European Commission's fiscal sustainability indicators and their use in economic policy coordination. The indicators root in the Commission's need to evaluate member states' compliance with EU fiscal rules to make sure that member states meet the public debt target of 60% of GDP. The first generation of these indicators measured projected shortfalls in revenues and debt repayments necessary to achieve the debt target by 2050. Yet, after the 2008 financial crisis, an additional indicator was developed in order to integrate an "early warning" mechanism for potential fiscal risks. Berti argues that these indicators, which place all member states on a spectrum from more to less sustainable fiscal positions, are used to provide input into the "country-specific recommendations" on fiscal sustainability, adopted by the Commission every year for each member state. In this understanding, the indicators seem very powerful, but Berti maintains that their role is not to determine or pre-empt policy but to support sound policy advice that is formulated by the Commission. She rejects the idea that the indicators become targets amounting to the monitoring and enforcement of the fiscal rules of the Stability and Growth Pact.

Alyssa Dougherty, Amy Gryskiewicz, and Alejandro Ponce discuss the methods used for developing the World Justice Project's Rule of Law Index, an annual index that represents a comparative measure of the rule of law for 113 countries in Chap. 11. They explain the challenge of developing an aggregatable measure of a concept that is inherently contestable—reflecting on the

inevitable normative choices that had to be made to try and quantify the idea of the rule of law. The authors explain the proxies they identified as components of the rule of law, and the methodological mix of expert questionnaires and public opinion measures used to develop the index. The method represents a complex and resource-intensive endeavour, requiring 113 opinion polls and the involvement of more than 1000 experts in reporting. The underlying premise—not much different to that of the World Bank's World Governance Indicators—is that the expansion of data points and inclusion of different types of data will (with weighting, normalisation, and aggregation) enable the index to capture the underlying quality of the rule of law in a given country (Urueña 2015).

In Chap. 12, Hazel Feigenblatt and Johannes Tonn (former and current directors of research at NGO Global Integrity) give an account of how Global Integrity's approach to indicator production has evolved over time. Their initial production of the (highly visible and successful) index of government "integrity" was followed by a decision to stop ranking countries, on the grounds that it proved not to be an effective policy-making tool. Interestingly, Global Integrity also eventually concluded that neither the provision of data nor the data itself was sufficient to impact on governance, and they sought to re-conceptualise the role of external best-practice benchmarks in motivating policy reform. The authors describe how Global Integrity instead developed contextually relevant "best-fit" approaches and benchmarks to evaluate performance and institutional reform strategies, even if this meant giving up on cross-country comparability.

The chapters by Rotberg (Chap. 2); Ivanyna and Shah (Chap. 18); Obinyeluaku (Chap. 13); and Raghunandan (Chap. 17) can be classified as written by authors closely associated with an indicator or measure that have contributed to its production.

Robert I. Rotberg criticises existing measures of good governance in Chap. 2 for their heavy reliance on subjective perception measures and maintains that objective measures can be derived from the outcomes of governance. He argues that outcome measures best capture whether or not a government is "performing" well in its delivery of essential public goods such as safety, educational opportunity, and economic opportunity. He acknowledges criticisms as to whether even these "objective" outcomes can be effectively measured and makes transparent his normative theory as to why good governance should be understood as the effectiveness of government in delivering these public goods to citizens.

Chapter 18 by Maksym Ivanyna and Anwar Shah adopts a more metrological stance in relation to existing governance indicators, arguing that citizen

perception is an accurate measure of governance quality. They use survey data from a range of data sets, extracting data concerning responses to those questions that they argue are relevant to governance quality. These include data on answers to questions such as "how much confidence do you have in the police?" or "would you fight for your country?" Citizen perception-related data is, thus, taken as a proxy for broad dimensions such as "accountable governance" and "responsive governance". Once again, a theoretical account of what good governance is must mediate the relationship between the data assembled and the dimensions of governance it is said to measure. Nonetheless, the authors argue that these measures capture something about the reality of the experience of governance within countries that otherwise is not captured.

In Chap. 13, Moses Obinyeluaku examines a case study in which an indicator becomes a target in order to depoliticise resource allocation decision-making. He describes the creation of an oil-price-based fiscal rule in Nigeria, which attempted to constrain government spending by establishing a reference or benchmark oil price. The benchmark price was set by a model that established a long-run average price by imitating a ten-year average oil price. The overall ambition of turning this indicator into a target for budgeting purposes was to restrain the procyclicality of actual changes in the oil price on the Nigerian budget. The chapter ultimately concludes that the process of setting the benchmark oil price itself became politicised, which undermined to some extent the effectiveness of this measure in restraining the procyclicality of oil price changes on the budget.

Chapter 17 by T.R. Raghunandan explores recent Indian experience with trying to measure and document corruption and notes that the most successful efforts have tended to be sector specific and citizen driven. Rather than to produce indexes and aggregate quantitative measures, civil society groups have used "social audits" of public expenditure in certain sectors to identify graft and wastage of resources. He also describes the "crowd-sourced" corruption-reporting platform, ipaidabribe.com (IPAB). The website generated publicity and effectively gathered indicative but unrepresentative data on corruption incidents and patterns derived from a collection of "citizens' reports". Raghunandan notes that, as alternatives to quantitative indicators, investigative and auditing approaches to measuring corruption had given civil society useful information. Comparisons between states had put the spotlight on those states which performed poorly. He notes however that, while the global corruption rankings provoke debate and controversy, the absence of national-level quantitative measures impeded civil society's ability to campaign for particular reforms at the national level and to track change over time.

The chapters by Cooley (Chap. 3); Alonso and Garcimartin (Chap. 4); Malito (Chap. 5); Burger (Chap. 6); Greenwood (Chap. 14); Musaraj (Chap. 15); Urueña (Chap. 16); Morse (Chap. 19); Blot (Chap. 20); and Merry (Chap. 21) are contributions by academics from diverse disciplinary perspectives, including political science, economics, law, history, philosophy, anthropology, and environmental studies.

In Chap. 3, Alexander Cooley argues that global corruption rankings adopt an inaccurately nation-state-centred perspective for understanding factors that enable grand corruption. Such indicators misdiagnose the essential facilitating causes of large-scale corruption, as they ignore the extent to which hidden wealth is dependent on a global network maintained by laws and regulations of "good governing" states. Thus, global corruption indexes "nationalise" the phenomenon of corruption when in fact it rests on transnational foundations ignored by such indicators. The states that facilitate grand corruption are often those which score well on national measures of corruption.

José Antonio Alonso and Carlos Garcimartín argue in Chap. 4 that the current measures of good governance reflect a strong set of theoretical claims derived from institutionalist economics concerning the importance of protecting private property from the state. These theoretical biases are even increased by serious problems of data. The result is a set of measures that generate a misleading impression that single-best institutional frameworks for governance exist. They propose as an alternative a more contextual construction of measures of institutional quality that is both country and time specific. In addition to quantitative measures, such an alternative approach required engagement with a broader range of social science insights in order to reconstruct the context in which the institutional performance is evaluated.

Debora Valentina Malito develops in Chap. 5 an extended critique of the presuppositions of measures of governance and stateness. She comprehensively reviews the historical emergence of such measures and explores the many conceptual and empirical problems they encounter. The chapter provides an essential roadmap and overview of the complex recent genealogy of measurements of governance and the state, and a trenchant analysis of its normative baggage.

In Chap. 6, Paul Burger underlines the inherent contestability of the concept of fiscal sustainability, a definitional complexity, which is further complicated by any attempt to define fiscal sustainability in light of a broader concept of sustainable development. He asks how sustainability could be rendered amenable to benchmarking or quantification in light of these definitional issues and argues that the inherent plurality of the concept of sustainability can only appear to be overcome through indicators. Where values conflict, as

they will in attempts to realise a multidimensional concept such as sustainability, choices must be justified through contextual arguments. He concludes that any one-dimensional sustainability assessment should be mistrusted, including one-dimensional assessments of fiscal sustainability.

In a similar vein, Daphne T. Greenwood criticises the understanding of fiscal sustainability as a single debt to GNI ratio in Chap. 14. Commonly used definitions of fiscal sustainability rested on a particular, neo-liberal theory of the economy and of the sources of macroeconomic stabilisation. Greenwood argues for an alternative understanding based on sustainable development theory, which prices not only private capital assets but human, social, and environmental assets. She proposes these dimensions of fiscal sustainability to be captured by local indicators that have been used at sub-national levels to capture important dimensions of sustainability which are ignored or rendered irrelevant through the use of single debt to national income measures.

Chapter 20 by Christophe Blot provides a critical analysis of the most common fiscal sustainability indicators, claiming that simple easy-to-compute and immediately available indicators might also provide misleading policy orientations. He claims that although indicators aim to provide simple and clear messages about the situation of public finances, they should not be considered independently from other economic and social variables. Many indicators of fiscal consolidation relied on fragile assumptions regarding the interest rate and the long-term growth. Main macroeconomic interdependencies and trade-offs usually not taken into consideration reveal that many indicators are derived from partial equilibrium settings that have a limited cognitive function on the status of public finance.

In Chap. 15, Smoki Musaraj uses anthropological methods to follow an indicator into the field with her case study of the impact of corruption indicators in Albania. She shows how a global knowledge form—a survey of corruption funded by the United States Agency for International Development and executed by a local institute—became enrolled in networks of local and international actors as an important reference frame for an unfolding political conflict. The impact of the corruption measure on local politics was indirect but also unanticipated by the funders and producers of the instrument.

Chapter 16 by René Urueña gives an account of the history of Transparency International's Corruption Perception Index, and reviews the many criticisms of both its conceptual foundations and its empirical basis. Urueña explores how the indicator's production and promotion became associated with a wider set of prescriptions not just concerning corruption but also economic liberalisation and other institutional markers of "good governance".

Stephen Morse reflects in Chap. 19 on the factors that may affect the uptake and use of an indicator, in particular indicators of sustainability. He

observes that there has emerged a complex ecology of sustainability indicators, driven in part by efforts to bridge the gap between technical specialists' research on sustainability and political decision-makers. Academic studies on the use of indicators suggest that the quality of an indicator is not a sufficient condition for its actual use in the policy process, and that the reasons behind the uptake on an indicator in a policy process may be exogenous to the indicator itself. Morse also considers the uptake of sustainability indicators by the media and finds that those which are most successful in terms of media salience are arguably those most redundant in terms of complexity as well as those that fit and tend to reinforce a pre-existing understanding of a problem (such as the debt-to-GDP ratio).

In Chap. 21, Sally Merry brings together numerous theoretical insights on the ways in which indicators work in contemporary global governance. Particularly noteworthy and helpful is her distinction between the "knowledge effects" of indicators and their "governance effects". The knowledge effects of indicators create a world that seems cognisable and comprehensible through the numerical measures. The governance effects engender repertoires for action and intervention shaped through the knowledge effects. The wide-ranging chapter provides a useful frame through which to reflect upon the preceding chapters of the Handbook and to re-assess the complex effects of quantification and its relationship to knowledge and governance in this period of globalisation.

In the concluding chapter (Chap. 22), the editors reflect on the three main topics discussed in this Handbook: a growing post-metrological trend, the relevance of normativity, and the complex and contextual power of indicators. They conclude that scholars and practitioners agree on the fact that measuring matters because of its instrumentality in governing. Moreover, all contributors were responsive to ongoing criticisms, even if they came to contrasting conclusions about the way forward.

The present Handbook takes stock of this plethora of conceptual issues regarding the use of indicators in global governance and reflects upon the variety of discordant interpretations surrounding them.

Notes

1. Poovey notes that "liberal governmentality" depended on self-rule rather than rule by coercion. Administering self-rule in a market society "involved understanding human motivations ... [and] as a consequence, the knowledge that increasingly seemed essential to liberal governmentality was the kind cultivated by moral philosophers".

2. For example, the now accepted notion of "national income" comparable across vastly different territories and peoples was once considered by economists to be an implausible object of measurement, susceptible to too many problems of data and conceptual validity to be tractable. As a result, the meaning of the numbers was closely contested, and the many judgements involved in quantifying national income were highly controversial until the matter became "settled" through its regular production and use in policy discussion—despite no real answer being provided to the scientific doubts about the numbers.
3. Desrosières refers to this as "accounting realism". Those without detailed knowledge of the way in which the number is made could easily slip into metrological realism—mistaking the number as a measure of a real thing in the world.
4. Broome and Quirk refer to "global benchmarking" as "an umbrella term for a wide range of comparative evaluation techniques that systematically assess the performance of actors, populations, or institutions. This can include techniques such as audits, rankings, indicators, indexes, baselines, and targets, all of which work on the basis of standardised measurements, metrics and rankings" (p. 815). They go on to argue 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" (p. 829).
5. The sources of this authority could be the power of the state producing the indicator (such as in the case of the United States' Trafficking in Persons' "Tier"-based ratings) or its incorporation into a form of conditionality, such as the United States' Millennium Challenge Corporation (MCC), which uses certain third party-generated composite indices as benchmarks to evaluate whether a state is eligible for some kinds of development funding.
6. "[Q]uantifying social phenomena requires translating things understood in idiosyncratic, systemic, or situational terms into things that can be counted. In making them commensurable, they must be viewed as, in some ways, the same, pulled away from their embeddedness in a holistic cultural and political context. Some features must be considered; others must be ignored."
7. The term "neo-liberal" has diverse meanings and connotations and may not denote exactly the same thing in its various uses. Here, we use it to refer to the political and economic policy agenda that gained strength in the late 1970s, which insisted on the rationalisation of public expenditures, the reduction of state regulation of economic activity, and the expansion of market dynamics in decisions concerning the allocation of public and private resources.

References

Akech, M. (2015). Evaluating the Impact of Corruption Indicators on Governance Discourses in Kenya. In S. E. Merry, K. E. Davis, & B. Kingsbury (Eds.), *The Quiet Power of Indicators: Measuring Governance, Corruption, and Rule of Law* (pp. 248–283). Cambridge: Cambridge University Press.

Bhuta, N. (2012). Governmentalizing Sovereignty: Indexes of State Fragility and the Calculability of Political Order. In K. Davis, A. Fisher, B. Kingsbury, & S. E. Merry (Eds.), *Governance by Indicators: Global Power through Quantifications and Rankings* (pp. 132–164). Oxford: Oxford University Press.

Bowker, G. C., & Leigh Star, S. (1999). *Sorting Things Out: Classification and Its Consequences*. Cambridge, MA: MIT Press.

Bradley, C. (2015). International Organizations and the Production of Indicators: The Case of Freedom House. In S. E. Merry, K. E. Davis, & B. Kingsbury (Eds.), *The Quiet Power of Indicators: Measuring Governance, Corruption, and Rule of Law* (pp. 27–74). Cambridge: Cambridge University Press.

Broome, A., & Quirk, J. (2015a). The Politics of Numbers: The Normative Agendas of Global Benchmarking. *Review of International Studies, 41*(5), 813–818.

Broome, A., & Quirk, J. (2015b). Governing the World at a Distance: The Practice of Global Benchmarking. *Review of International Studies, 41*(5), 819–841.

Bulmer, S., & Padgett, S. (2004). Policy Transfer in the European Union: An Institutionalist Perspective. *British Journal of Political Science, 35*(1), 103–126.

Bukovansky, M. (2015). *Corruption Rankings: Constructing and Contesting the Global Anti-Corruption Agenda*. In Cooley and Snyder 2015 (pp. 60–84).

de Burca, G. (2010). New Governance and Experimentalism: An Introduction. *Wisconsin Law Review, 2*, 227–238.

de Burca, G., Keohane, R., & Sabel, C. (2014). Global Experimentalist Governance. *British Journal of Political Science, 44*(3), 477–486.

Collier, S. (2012). Neo-liberalism as Big Leviathan or …? A Response to Wacquant and Hilgers. *Social Anthropology, 20*(2), 186–195.

Cooley, A. (2015). The Emerging Politics of International Rankings and Ratings: A Framework for Analysis. In A. Cooley & J. Snyder (Eds.), *Ranking the World: Grading States as a Tool of Global Governance* (pp. 1–38). Cambridge: Cambridge University Press.

Cooley, A., & Snyder, J. (Eds.). (2015). *Ranking the World: Grading States as a Tool of Global Governance*. Cambridge: Cambridge University Press.

Davis, K. E., Kingsbury, B., & Merry, S. E. (2012). Indicators as a Technology of Global Governance. *Law and Society Review, 46*(1), 71–104.

Desrosières, A. (1998). *The Politics of Large Numbers: A History of Statistical Reasoning*. Cambridge, MA: Harvard University Press.

Desrosières, A. (2001). How Real Is Statistics? Four Possible Attitudes. *Social Research, 68*(2), 339–355.

Desrosières, A. (2015). Retroaction: How Indicators Feed Back onto Quantified Actors. In R. Rottenburg, S. E. Merry, S. J. Park, & J. Mugler (Eds.), *The World of Indicators: The Making of Governmental Knowledge through Quantification* (pp. 329–353). Cambridge: Cambridge University Press.

Dutta, N. K. (2015). Tradeoffs in Accountability: Conditionality Processes in the European Union and Millennium Challenge Corporation. In S. E. Merry, K. E. Davis, & B. Kingsbury (Eds.), *The Quiet Power of Indicators: Measuring Governance, Corruption, and Rule of Law* (pp. 156–196). Cambridge: Cambridge University Press.

Espeland, W. (2015). Narrating Numbers. In R. Rottenburg, S. E. Merry, S. J. Park, & J. Mugler (Eds.), *The World of Indicators: The Making of Governmental Knowledge through Quantification* (pp. 56–75). Cambridge: Cambridge University Press.

Espeland, W., & Stevens, M. L. (1998). Commensuration as a Social Process. *American Journal of Sociology, 24*(1), 313–343.

Espeland, W., & Stevens, M. L. (2008). A Sociology of Quantification. *European Journal of Sociology, 49*(3), 401–436.

Fioramonti, L. (2014). *How Numbers Rule the World: The Use and Abuse of Statistics in Global Politics*. London/New York: Zed Books Ltd..

Foucault, M. (2007). *Security, Territory, Population: Lectures at the Collège de France 1977–1978*. Basingstoke/New York: Palgrave Macmillan.

Fukuda Parr, S. (2003). The Human Development Paradigm: Operationalizing Sen's Ideas on Capabilities. *Feminist Economics, 9*(2–3), 301–317.

Hacking, I. (1982). Biopower and the Avalanche of Printed Numbers. *Humanities in Society, 5*, 279–295.

Hacking, I. (1990). *The Taming of Chance*. Cambridge: Cambridge University Press.

Hacking, I. (2002). *Historical Ontology*. Cambridge, MA: Harvard University Press.

Hurrell, A. (2007). *On Global Order: Power, Values and the Constitution of International Society*. New York: Oxford University Press.

Kelley, J. G. (2017). *Scorecard Diplomacy: Grading States to Influence their Reputation and Behaviour*. Cambridge: Cambridge University Press.

Kelley, J. G., & Simmons, B. A. (2014). *The Power of Performance Indicators: Rankings, Ratings and Reactivity in International Relations (American Political Science Association Annual Meeting Paper)*. Washington, DC: American Political Science Association. Retrieved from https://ssrn.com/abstract=2451319.

Knill, C. (2005). Introduction: Cross-national Policy Convergence: Concepts, Approaches and Explanatory Factors. *Journal of European Public Policy, 12*(5), 764–774.

Lindenfeld, D. (1997). *The Practical Imagination: The German Sciences of State in the Nineteenth Century*. Chicago: University of Chicago Press.

Merry, S. E. (2016). *The Seductions of Quantification: Measuring Human Rights, Gender Violence and Sex Trafficking*. Chicago: Chicago University Press.

Merry, S. E., & Wood, S. (2015). Quantification and the Paradox of Measurement: Translating Children's Rights in Tanzania. *Current Anthropology., 56*(2), 205–217.

Merry, S. E., Davis, K. E., & Kingsbury, B. (Eds.). (2015). *The Quiet Power of Indicators: Measuring Governance, Corruption, and Rule of Law*. Cambridge: Cambridge University Press.

Poovey, M. (1998). *A History of the Modern Fact: Problems of Knowledge in the Sciences of Wealth and Society*. Chicago: Chicago University Press.

Porter, T. (1995). *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*. Princeton: Princeton University Press.

Porter, T. (2015). The Flight of the Indicator. In R. Rottenburg, S. E. Merry, S. J. Park, & J. Mugler (Eds.), *The World of Indicators: The Making of Governmental Knowledge through Quantification* (pp. 34–55). Cambridge: Cambridge University Press.

Power, M. (1997). *The Audit Society: Rituals of Verification*. Oxford: Oxford University Press.

Roberts, A. (2010). *The Logic of Discipline: Global Capitalism and the Architecture of Government*. New York: Oxford University Press.

Rottenburg, R., Merry, S. E., Park, S. J., & Mugler, J. (Eds.). (2015). *The World of Indicators: The Making of Governmental Knowledge through Quantification*. Cambridge: Cambridge University Press.

Sassen, S. (2007). *A Sociology of Globalization*. New York: WW Norton.

Schueth, S. (2015). Winning the Rankings Game: The Republic of Georgia, USAID and the Doing Business Project. In A. Cooley & J. Snyder (Eds.), *Ranking the World: Grading States as a Tool of Global Governance* (pp. 151–177). Cambridge: Cambridge University Press.

Serban, M. (2015). Rule of Law Indicators as a Technology of Power in Romania. In S. E. Merry, K. E. Davis, & B. Kingsbury (Eds.), *The Quiet Power of Indicators: Measuring Governance, Corruption, and Rule of Law* (pp. 199–221). Cambridge: Cambridge University Press.

Speich, D. (2011). The Use of Global Abstractions: National Income Accounting in the Period of Imperial Decline. *Journal of Global History, 6*(1), 7–28.

Strathern, M. (Ed.). (2000). *Audit Cultures: Anthropological Studies in Accountability, Ethics and the Academy*. London: Routledge.

Tribe, K. (1995). *Strategies of Economic Order: German Economic Discourse, 1750–1950*. Cambridge: Cambridge University Press.

Umbach, G. (2009). *Intent and Reality of the European Employment Strategy: Europeanisation of National Employment Policies and Policy-making?* Baden-Baden: Nomos.

Umbach, G. (2017). Measuring (Global) Governance—The Potential, the Practical and the Problematic Assessment of Governance Within and Beyond the State. In A. Triandafyllidou (Ed.), *Global Governance from Regional Cultural Perspectives: A Critical View* (pp. 45–69). Oxford: Oxford University Press.

Urueña, R. (2015). Indicators and the Rule of Law: A Case Study of the Rule of Law Index. In S. E. Merry, K. E. Davis, & B. Kingsbury (Eds.), *The Quiet Power of*

Indicators: Measuring Governance, Corruption, and Rule of Law (pp. 75–102). Cambridge: Cambridge University Press.

Voß, J. P., & Freeman, R. (2016). Introduction: Knowing Governance. In J. P. Voß & R. Freeman (Eds.), *Knowing Governance: The Epistemic Construction of Political Order* (pp. 1–33). Basingstoke/New York: Palgrave Macmillan.

World Commission on the Social Dimension of Globalization. (2004). *A Fair Globalization: Creating Opportunities for All*. Geneva: International Labour Organization. Retrieved from http://www.ilo.org/public/english/wcsdg/docs/report.pdf.

Part I

Conceptualising and Contextualising Indicators as Instruments of (Global) Governance

2

Good Governance: Measuring the Performance of Governments

Robert I. Rotberg

Introduction

Perfecting the instruments of world order and improving outcomes for the globe's citizens depend upon the actions and advocacies of presidents, prime ministers, provincial governors, and mayors—political leaders of nation-states, provinces, and cities—and the extent to which each practises rudimentary good governance, the better to produce prosperity and peace within their jurisdictional boundaries.

Robustly governed peoples, or peoples who have achieved reasonably high levels of governance, are wealthier, avoid civil conflict more readily, and obtain abundant social returns. Good governance and stability, moreover, go hand in hand, as do all kinds of positive developmental indicators. Additionally, good governance attracts foreign investment, helps to reduce poverty, enhances educational attainment, and gives citizens a solid sense of belonging to an uplifting common enterprise. Together, these are some of the beneficial results of good governance.

But what is "governance"? Is it a fuzzy or generic concept that cannot be fully defined? Or can its content and its essential meaning be fully specified in order to strengthen the ways in which both the practical aspects of governance and its deeper existential attributes are helpful in building theory, enhancing

R.I. Rotberg (✉)
Harvard Kennedy School's Program on Intrastate Conflict, Cambridge, MA, USA

D.V. Malito et al. (eds.), *The Palgrave Handbook of Indicators in Global Governance*,
https://doi.org/10.1007/978-3-319-62707-6_2

understanding of what governments do and how they operate as states and regimes, and providing realistic options capable of framing and driving policy?

There is virtue in being as precise as possible about what governance is and what it is not. In at least this case, precision and specificity materially assist in framing policy options and making policy decisions. We should also attempt to show how governance can be systematically measured using sophisticated quantitative criteria that employ objective rather than subjective data. The methods of measurement need to be replicable and genuinely informative (Rotberg 2004b).

Governance as the term used, defined, and referred to in this chapter concerns domestic governance, that is, governance within a domestic governmental jurisdiction (nation, province, or municipality). The word has a distinctly different meaning and is understood in a much more general way when it is linked with the term "international" or "global", as in "global governance" or "governing the world system". Biersteker has produced the most comprehensive definition of global governance. He calls it a "purposive" system of rules "that operate at the global level". Furthermore, global governance entails both "decisions that shape and define expectations at the global level" and a social relationship between "some governing authority and some relevant population or public that recognizes and acknowledges that authority as… [legitimate]" (Biersteker 2015, p. 158).

Domestic governance is an entirely different animal since it concerns how individual national governments operate, not how these governments function within an overarching global system. Rothstein equates good governance at the domestic level with impartiality—with a refusal of governing institutions and their functionaries to have favourites in the exercise of public authority (Rothstein 2011). Dahl proclaims "political equality" to be the essence of good governance, while Birdsall sees it as "fair competition" (Birdsall 2016). Fukuyama (2013a) articulates a notion of governance which emphasises bureaucratic capabilities as a key aspect of any definition. Kaufmann et al. (2010, p. 4) define good governance as the "traditions and institutions by which authority in a country is exercised". The United Nations Development Programme (UNDP) calls governance a "system of values, policies and institutions by which a society manages its economic and social affairs […] It is the way a society organizes itself to make and implement decisions" (UNDP 2000).

Rothstein (2011) seeks normatively and procedurally to measure a state's "impartiality" as a proxy for the overall quality of a government: "Just political procedures are those that by and large can be seen as impartial by groups with very different conceptions of 'the good.'" The Quality of Government Institute

at the University of Gothenburg, which Rothstein led, surveys experts' opinions (1000 in 126 countries) to estimate "impartiality" and rank countries according to quality—how good their governments are (Rothstein 2011, p. 12–23; Rothstein and Teorell 2013). But impartiality may or may not carry with it the ability to deliver—that is, to perform.

Fukuyama (2013b, p. 3–4) prefers to define governance as a government's ability "to make and enforce rules, and to deliver services", whether within a democratic framework or not. For him, "governance is […] execution" (Fukuyama 2013b, p. 3–4). How a regime administers itself is critical. He builds on Max Weber's criteria for successful bureaucracies: technocrats are selected and motivated by merit, remunerated fairly, and subject to discipline and control.

My work on governance (here and earlier), in contrast, has been predicated on governance—and hence good governance—being defined as the performance of national, provincial, municipal, and other governments in delivering specific articulated political goods.[1] The designation of such political goods is much more descriptive than normative; the political goods in question are those that citizens everywhere—in every culture, in every political jurisdiction—expect their governments (and their political leaders) to provide. As a result, governance becomes the effective or ineffective distribution of expected services to constituents, the inhabitants, or citizens of whatever political jurisdiction (Rotberg 2007).

Governance is thus not an approach, a method, a set of competencies, or a normative design. It is a collection of responsibilities, the delivery of which can be measured directly or by proxy. Without the check of measurement, "governance" descends too readily into attitudes and generalities, as in survey assessments calibrating whether or not Ruritania is "impartial" or is effectively bureaucratic.

Governance and Democracy

Fortunately, this conception of governance neatly replaces "democracy" when, as scholars or practitioners, we seek to estimate the extent to which a particular regime is or is not governing its people well. Whereas "democracy" as a concept has many contested definitions, and whereas the extent to which a political entity is or is not "democratic" over decades and across diverse continents and chronological periods has proven almost impossible to specify in a manner that invites widespread agreement, quantities and qualities of "governance" and whether a polity exhibits good or bad governance, can be

decided on the basis of neutral data (Rotberg 2012). The nature of Ruritania's "governance" hence depends on specific kinds of service deliveries that can be derived from readily available statistical bases. Doing so does not need to rely on the very imprecise, tripartite, and somewhat arbitrary categorisation of the nations of the world as "free", "partly free", or "unfree", or as "strong", "weak", "failed", or "collapsed" (Rotberg 2004a, 2017).

It is much easier and more satisfactory to measure governance results than to try to measure democracy outcomes. To compare policy purposes and policy-option decisions—how citizens are faring under regime X or Y—we need to compare governance (not democracy) in X and Y. Governance is indeed a construct that enables us exactly to discern whether citizens are progressing in meeting life's goals. Are they better off economically, socially, and politically than they were in an earlier decade? Are their various human causes advancing? Are their governments treating them well, attempting to respond to their various needs and aspirations, and relieving them of anxiety? Simply comparing national GDPs, life expectancies, or literacy rates provides some helpful distinguishing data (as in the Human Development Index), but governance data are more comprehensive, telling, and useful. Assessing governance tells us more about actual results in various developing societies than we would learn by weighing the varieties of democracy in such places (Rotberg 2013).

Exporting democracy from the North to the South, from developed to less-developed nations, is often a stated goal, but what world powers and the nations of the developed world seek to transfer is more than "democratic practices"—it is the whole bundle of political goods that in fact amount to good governance. Democracy is both a narrower and a more abstract notion than "governance". Nation-states that embody good governance perform effectively for their citizens. They deliver abundant quantities and high qualities of the essential political goods that comprise governance.

Political Goods and Measuring Governance

The performance of governments means the delivery of five bundles of political goods (divided into 57 underlying sub-categories) that citizens in any kind of political jurisdiction demand and expect (Rotberg and Gisselquist 2007, 2008, 2009). Those five bundles are Security and Safety, Rule of Law and Transparency, Political Participation and Respect for Human Rights, Sustainable Economic Opportunity, and Human Development (schooling and health).

Each of these bundles or categories of political good can be measured by analysing and aggregating the results of the sub-categories that inform an assessment of governance within a political jurisdiction. The sub-categories directly facilitate measurement, as in annual homicides per 100,000 of the population for Safety or number of deaths in civil conflict per capita for Security. Likewise, to assist in compiling results for Sustainable Economic Opportunity, it is possible to provide concrete measurements of the length of the paved road network in a country or the number of mobile telephones per 100,000 of the population. GDP per capita numbers also help to develop a score in the Economic Sustainability category.

For Human Development numbers, we can employ several readily available educational and health statistics. However, in some cases, proxy results must be substituted for direct national statistics. For example, overall life expectancy and maternal mortality numbers help us to estimate the performance of a government in terms of its ability to provide better health outcomes. Similarly, the total availability of potable water to citizens assists our understanding of how well a developing country is being governed—how well it is satisfying the essential requirements of its population.

Most of the available data on which the measurements are based are nationally supplied and nationally developed. That, of course, is a weakness that to some extent undermines the ultimate results of any attempt to specify amounts of governance at the national level, especially across the developing world (Jerven 2013). One safeguard is that all of the results and rankings are expressed comparatively, against peers. Thus, even if the underlying data are in some cases poor, one of the objects of estimating governance for diagnostic and policy purposes is to compare Ruritania to XRuritania, not to develop a precise chart of absolute performance.

Fortunately, too, many of the nationally supplied databases, especially in the health sector, are regularly massaged and normed according to internationally acceptable statistical methods. It is also possible directly to collect and measure homicide or road mile (or other) numbers without relying on weak national coordinating or statistical agencies.

A few governance measures are also available globally from tried and trusted sources, such as Transparency International's Corruption Perceptions Index (for transparency) and the World Justice Project's Rule of Law Index for parts of the rule of law. Election results (whether free or not) help to supply numbers for Political Participation, and several well-developed human rights compilations (based on data derived from diplomatic inspections and surveys) offer helpful scorings for Respect for Human Rights (Rotberg and Bhushan 2015).

If the aim of developing and developed world leaders and of those in the scholarly and humanitarian communities is to help improve life outcomes and raise the efficacy of governmental decision-making, then the kinds of detailed measurement initiatives outlined here are essential. They provide critical feedback about whether or not governments are doing their jobs.

Governance scoring methodologies, as set out here and elsewhere and as employed formally by the original Index of African Governance and informally in other indexing schemes, offer detailed approximations of actual service delivery successes and failures (governance). They substitute quantifications for impressionistic or opinion-based ratings that report on the *perceived* quality of governmental operations, their *perceived* impartiality, the extent of a government's *perceived* bureaucratic autonomy, or its *perceived* capacity to coax or coerce citizens. Only by closely examining actual governance results based on statistics and proxies, and creating ways to do so in a numerically sound manner, can we distinguish the governments that are producing abundant political goods (good governance) from those that no longer are or never have done.

To answer critics of this predominantly objective method, governance outcomes are best viewed as tangible, not subject to collections of observations. Governmental outcomes do not need to be judged by stylistic or artistic criteria. This more rigorous scheme, moreover, makes epistemological and parsimonious sense. It is tidy and transparent and capable of being tested and replicated (Rotberg 2014, 2015).

Inputs or Outputs? Quality or Capacity?

Much of the other, dissenting, work on measuring governance has found it easier and more satisfying to examine inputs rather than outputs, and has done so subjectively by estimating bureaucratic and other "capacity" or by assessing budgetary procedures, styles of financial management, or the ineffable "quality" of governments. Rothstein's wise focus on "impartiality", for example, may not carry with it the ability of a government to deliver—that is, to perform. Degrees of "impartiality" are also very hard fully to measure: estimating amounts of "impartiality" may depend too much on the observations of survey respondents—even experienced and well-meaning "experts"—whose findings are inevitably tainted by the possibility of selection bias.

There are innumerable quality studies of governance based on the opinions of experts recruited especially for the task or on surveys of citizen satisfaction. Among these helpful and otherwise exemplary efforts are those undertaken by

the Varieties of Democracy Project and indexes such as the Bertelsmann Stiftung's Transformation Index. Freedom House's *Freedom in the World*, another long-established and worthy examination of democracy (and, implicitly, governance), annually declares almost every global polity to be either free, partly free, or not free based on the opinions of experts. Nevertheless, its scoring, like that of Bertelsmann, is inherently subjective, with abundant opportunities for selection bias (Freedom House 2013). The UNDP's way of defining governance (above) is also very general and subjective. Attempting to conceptualise governance in such a broad all-encompassing manner—as many indexes and similar efforts prefer—becomes largely an exercise in subjective speculation.

Bratton (2013a), one of the founders of the Afrobarometer, suggests that "governance is the act or process of imparting direction and coordination to governmental organisations in an environment". His definition of governance is close to Fukuyama's, as it is "administrative and economic" as well as political. Bratton believes in the utility for measuring purposes of large-scale social surveys, as they can indicate citizen or consumer satisfaction. An example he cites from the administrative sphere is the *perception* of a national leader's observance of the rule of law as discerned across countries from responses to Afrobarometer questions (Bratton 2013a).

The World Bank's annual compilation, Worldwide Governance Indicators (WGI), is widely used. It measures quality of national governance by aggregating other indexes of governmental effectiveness, regulatory quality, stability, and control of corruption—all attributes capable of being estimated by crowd sourcing (surveys of experts), but more difficult to calibrate using nationally generated statistics. The indicators used are largely normative, encompassing policy preferences rather than measuring the satisfaction of citizen-requested priorities (as in the output-oriented index proposed here). Rothstein and Teorell (2013) consider the WGI's definition of governance as too broad, especially its normative emphasis on "sound policies". They further explain that the Bank's emphasis on the input side of the governance equation "makes it impossible" to provide true results for governmental performance (Rothstein and Teorell 2013, p. 3; see also Kaufmann et al. 2010; Gisselquist 2015).

As Fukuyama (2013a) indicates, expert surveys are inherently weak. Unless the experts have a common notion of "governance" or "regulatory effectiveness", each may answer the questions posed honestly but from vastly different perspectives. One person's corruption, in other words, may be another's reciprocal "gift-giving". Rule of law, he says, may even mean one thing in one region and something very different in another. Some may translate rule of

law as "property rights" and another as constitutional and other constraints on the executive (Fukuyama 2013a).

Arndt and Oman (2006) concur with Fukuyama and other critics opposed to basing index results and governance comparisons largely, if not entirely, on the perceptions of experts. Such views, they say, are "inherently subjective" and "non-replicable". Moreover, perception-based indicators often reflect the views of businessmen, and "users tend to rely on the same indicators which they see their peers using", so there is a "bubble effect" and "herd behaviour" (Arndt and Oman 2006, p. 90). Their report, along with work by Thomas (2010), includes lengthy critiques of the methodology employed by the WGI.

Inputs are more accessible and thus easier to measure. Fukuyama and many others who have produced existing scholarship on governance—and the makers of many governance indexes—often focus on inputs, not outputs, when they characterise governance. However, as Fukuyama agrees, inputs—"good procedures and strong capacity"—"are not ends in themselves" (Fukuyama 2013b, p. 11).

Outputs (results) are for what nation-states and governments (and thus governance) exist. Their role, and the role of the modern polity, is to produce (perform) on behalf of constituents. As Fukuyama rightly says, measuring outputs "could" provide "some idea as to how governments are performing". In his view, however, there are decisive drawbacks to the use of outputs. For example, improvements in the educational or health areas are not necessarily the consequences of governmental action (Fukuyama 2013b, p. 11). Such results could (as many economists also believe) flow from the contextual situation because of an existing resource base or be products of historical circumstances. Andrews' (2013, p. 5) reservation (with which I mostly disagree) is cautionary: indicators of governance "really reflect a nation's level of development" and not its governance. Fukuyama also worries that there are too many methodological problems with the measurement of many kinds of outcomes and that outcomes can be influenced too much by procedural inputs—how a regime delivers results could influence the type of outcome (Fukuyama 2013b).

Fukuyama proposes that outputs should be considered independent variables explained by state quality rather than be used as measures of capacity themselves. He further suggests that the quality of government (or governance) is to be found at the intersection between what he calls capacity and what Huntington (2006) describes as "bureaucratic autonomy"—the ability of bureaucrats to carry out the policies of a state according to broad guidelines and with little micromanagement. Capacity, imperfectly defined, includes the ability of a state to perform essential functions, such as (but not exclusively) the ability to extract taxes and obtain census information.

These kinds of caveats are all reasonable and worth pondering, but it is so much easier and complete directly to measure the quantity and quality of a government's service deliveries by scoring what its actions have accomplished (as set out here and as developed in the Index of African Governance).

Andrews' (2013) definition of governance leans in the general direction proposed in this chapter: that the core theoretical understanding of "governance" should be "the exercise of authority by governments on behalf of citizens". Governance indicators, he writes, should therefore focus (as the Index of African Governance does) on "specific fields of engagement" in which governments perform on behalf of citizens. "Indicators should emphasize outcomes […] the true indicators of governance" (Andrews 2013; Gisselquist 2012; Rotberg and Gisselquist 2009). As I argue, that is the cleanest and most rewarding approach to comparing Ruritania and XRuritania.

Relying on perceptions (as do Bratton and many others, even Freedom House) is misleading because of possible selection bias. If our goal is to understand and strengthen governance across cultures and continents, we do better to employ methods of measuring how governments perform that are as insulated from impressionism and anecdote as possible.

Subjectivity is dangerous. Polling such as Bratton recommends and on which many others rely—no matter how broad or how carefully representative—cannot objectively tell us how a government performs (Bratton 2013b). Polls and other collectivities of opinion merely tell us, and usefully, what citizens think or what they perceive—always depending on how well and how precisely the survey questions are posed.

Outcome Measures

If governance is indeed "performance"—the delivery of services, as posited at the outset of this chapter and as Fukuyama seems to agree—then arguably the most important measures of that delivery must be both the quality and the quantity of those services. This chapter offers a definition of governance that tries hard not to be prescriptive or normative. It proceeds from a summing of the needs, desires, and expectations of inhabitants of jurisdictions, usually citizens. What is it that citizens expect or demand of their governments? What is it since the seventeenth century that citizens have required of their monarchs and later of their states and nations? My definition is founded on a bottom-up method of defining governance that emphasises results.

If we agree that citizens (originally taxpayers) expect their governments to perform in such a manner that they will be secure (free from being invaded

or free from civil war and intrastate turmoil) and safe (free from crime and personal endangerment); if we agree that citizens desire something akin to the predictability and backing of a robust rule of law that delivers sanctity of contract as well as a fair adjudication of disputes between persons; if we agree that most inhabitants of most states prefer not to be cheated by corrupt practice; if we agree that citizens prefer to participate in rule setting and thus in governing themselves, or at least prefer to have a voice in agenda setting; if we agree that individuals prefer to prosper, that is, to eat more and better food, to be housed adequately, to be paid fairly for their labour, and to believe that they are free to use their own skills to better themselves; and, finally, if we agree that citizens generally look to states to provide educational opportunities, better healthcare, clean water, a minimally polluted environment, and so on, then it makes perfect sense to compare better and poorer ways in which these needs are realised nationally—thus better and poorer ways in which states perform for their taxpayers and inhabitants, or a composite of political good delivery that may conveniently be labelled "governance". The assumption, tested in the field and by various methods of polling opinions, is that the above expectations are desired everywhere. No one wants poor pay or less desirable schooling. No one anywhere wishes to be unsafe or insecure. No one abides corruption, even though many suffer from its pernicious grip.

The way to make these necessary comparisons—to calculate the manner in which the state uses its capacity and a lesser or greater sense of bureaucratic autonomy to satisfy its citizens—is to measure variables such as participation or educational opportunity. To measure the variables, it is essential to examine results. There is no better way of estimating how successfully a state has met its obligation to serve (perform) than carefully to calculate outcomes. It is possible, to be sure, to ask citizens if they are "satisfied" with a government's performance. That happens periodically through elections, and also through consumer surveys, which are helpful but hardly definitive. More exacting and more useful when one tries to compare disparate polities or attempts to diagnose how a nation-state could do more for its citizens—how it could perform better and more completely—are examinations of actual quantifiable results.

Since it is difficult to measure governance performance across the five categories of political goods directly, we sometimes use proxies (as explained above). Fukuyama (2013b, p. 9) suggests that services are hard to estimate, in part because some of the potential tests are unsatisfactory. He cites as bad choices the use of multiple choice examination scores to measure educational

outcomes or the rate of case clearances to measure the quality of justice or the rule of law. However, there are better proxies for those political goods, and across the 57 variables that comprise my preferred output measurement dataset, a number are very robust, a few are unhappily qualitative and subjective but have been long in use, and all are demonstrably helpful in estimating the performance of governments. That is, they do a very strong, albeit imperfect, job of measuring what we want to measure. They offer the kinds of hard data that truly enable us to compare countries to countries, provinces to provinces, municipalities to municipalities, and so on.

The other (largely subjective) method, as used by Freedom House, Bertelsmann, the Legatum Institute's Prosperity Index, the Fraser Institute's Economic Freedom Index, Save the Children's Mother's Index, the Stanford Food Research Institute's Hunger Index, and even the Happy Planet Index, produces helpful approximations of the realities of at least some selected aspects of governance among and between nations (for the indexes, see Rotberg and Bhushan 2015). However, their scores are based on opinions, feelings, codings, anecdotal understandings, and the like. For example, the widely respected World Economic Forum's Competitiveness Index depends on a survey of the views of a limited number of business executives within a country. Those kinds of index have great difficulty measuring any more than what observers and opinion makers *believe* is happening, not unchallengeable reality. Experts may say, for example, that country A has better educated citizens than country B, when an examination of real results—persistence levels, numbers of students who go on from secondary to tertiary education, and so on—might reveal that country B is in fact producing better educational outcomes than country A.

Methodologically, the notion of governance as developed and portrayed in this chapter is more definitively measured by outputs, not inputs. For example, in evaluating a national or local medical system, a country's healthcare budget matters far less than how that money has been used (or siphoned off) to improve health outcomes. Good ministerial attitudes and fine words matter far less than results. No amount of jawboning can obscure a lack of discernible results.

For example, when eschewing surveys of experts or other methods of eliciting public opinion as proxies for several of the otherwise hard-to-measure five major components of governance, it is wise to employ such internationally standard variables as life expectancy and maternal mortality rates for outcomes in the health arena. To help calibrate sustainable economic opportunity, it is possible to calculate a nation's GINI coefficient ratios, an indication

of equality, or to employ various IMF indicators to estimate how sound its money and banking systems are. For Safety, homicide rates (see above) are better than asking experts whether Ruritania is safe. For Security, it is more instructive to employ one of the many available international datasets that annually supply the number of casualties in civil warfare than it is to survey experts. For a number of the variables, and certainly for the general 5 categories, there are abundant proxy data contained in a variety of the 93 indexes, indicators, and databases explored in *On Governance* (Rotberg and Bhushan 2015).

Most useful statistics for such exercises are assembled internationally by such estimable establishments as the World Bank, the International Monetary Fund, the World Health Organization, the UNDP, and the UN Children's Fund. Admittedly, because nearly all of these numbers are ultimately derived from national statistical offices, and because many of these offices are underfunded and understaffed, the development of reliable data is sometimes compromised. This is a weakness of the objective output-oriented indexes that exist (see Jerven 2013), but the problems posed by selection bias and subjectivity are greater.

In Africa, it is counterintuitive that poor Malawi should be better governed than prosperous and bustling Kenya, but that is what the Index of African Governance rankings shows year after year, probably because Malawians demonstrate positive outcomes despite poverty and Kenyans fall short because of greater ethnic conflict, greater corruption, and poorer educational attainment. Kenya is a busy, populous, modern tourist destination and the industrial and transport hub of East Africa, whereas Malawi is a sliver of landlocked territory dependent almost exclusively on the sale of tobacco to China. But Malawi has a solid and reasonably independent court system in contrast to Kenya's compromised one. Kenya has experienced far more ethnic tension and violence than Malawi. Kenya's elections in 2007 were rigged and Malawi's in 2009 and 2014 were not, leaving serious questions about the integrity of Kenya's democracy and its respect for human rights. Hence, although Kenya's GDP per capita rates are much higher than Malawi's (and higher for that matter than those of Tanzania, Mozambique, and Uganda), Kenya's ratings for Security (even before the combat with al-Shabaab) and Safety, for Rule of Law and Transparency (corruption is more prevalent in Kenya than it is among its neighbours), and for Participation are lower than its peer group and competitors. Moreover, despite its greater wealth, Kenya's prosperity is less equally distributed than a number of its second-tier competitors. Its arteries of commerce, per head, are less robust than in comparable countries. Furthermore, despite its progress in educational and health matters, when compared to

other second-tier countries, Kenya ranks poorly on a per capita basis. Thus, appearances and anecdotal evidence turn out to be less impressive when weighing governance than results gathered and expressed quantitatively.

What Measuring Performance Provides

In the absence of a results-based method of weighing the performance of governments, drawing such distinctions between countries would be imprecise, even potentially inaccurate. Using a vague notion of "quality", a proxy for "capacity" or a proxy for "bureaucratic autonomy" would leave differences between countries and across regions that would be very hard to substantiate, even impressionistically. Policy makers want precise calibrations, not rough estimates. Heller (2013) proposes employing a very complicated array of "second-generation" data to do the job of measurement when much simpler, more direct ways of assessing good governance are easily available, as specified in this chapter.

When performance is measured, performance usually improves, especially when it is measured and the results are reported. Indeed, we know from anecdotal and some limited controlled experimentation that we can enhance the delivery of political goods to citizens and thus strengthen governmental performance by indexing relative accomplishments across a universe of diverse governmental tasks. Making these possibilities better known is one of the goals of this chapter.

In order to make the measurement of governance even more meaningful, and to make various kinds of indexes of governance maximally useful to policy makers, opinion shapers, civil societies, and donors, as well as to assist governments, the five overarching categories of political good already discussed must be further refined so that the full range of governance can be measured in its entire complexity. As suggested above, it is sensible to break down the 5 major categories of political goods into 57 or more indicators or variables, for example, 22 under Human Development and 5 under Safety and Security. Most of these variables can be assessed (as discussed above) by employing objective rather than subjective data.

Employing composite or specialised indexes and datasets enables observers and analysts to assess the state of governance across continents and within a region, country versus country, and not against arbitrary global standards. Countries can be scored on each variable and across several years. For example, within the Sustainable Economic Opportunity category, countries can be rated not only on the basis of their annual GDP per capita achievements but

also on their levels of inflation, the integrity of their banking systems ("contract-intensive money"), and the length and availability of robust arteries of commerce (such as mobile phones, roads, and railways).

The examination of all of these components of governance permits a full diagnostic portrait of each country. Hence, we can know with some significant precision how well and how poorly each part of the globe is governed. We can determine which aspects of governance in each country need urgent attention and which are in lesser need of remedial action. Furthermore, we can assemble a precise list of a region's outstanding governance issues. With such diagnostic information at hand, the political leaders, corporate moguls, and civil societies of that region and of the globe can at the very least come to appreciate what they must do to strengthen their individual national governances. At that point, they possess the evaluative tools with which to improve their governmental performances as they confront the consummate developmental and social challenges of the next decades.

Conclusion

This chapter has argued that precise measurements of governance as a whole, and of governance separated into its component parts, permit researchers and policy makers to separate good performers from bad performers, regionally and globally. Carefully and painstakingly measuring governance by the outcome method can show whether regimes are delivering necessary and desirable governmentally provided performance results to their citizens. This concept of governance also enables an existing government, civil society, or donors to appreciate which parts of an overall system are working well and which poorly. Critical decisions can thus be made that can, in the best of circumstances, improve living conditions for the inhabitants of the developing world.

Improved governance is the goal. Measurement according to the criteria set out in this chapter, followed by concerted action, is the best way to achieve that goal.

Notes

1. For earlier writings, see the Rotberg entries in the references.

References

Andrews, M. (2013). Rethinking Governance Indicators. *APSA-Comparative Politics Newsletter, 23*(*1*), 5–7. Retrieved from http://comparativenewsletter.com/files/archived_newsletters/2013_winter.pdf.

Arndt, C., & Oman, C. (2006). *Uses and Abuses of Governance Indicators*. Paris: OECD Development Centre.

Biersteker, T. (2015). Global Governance. In R. I. Rotberg (Ed.), *On Governance: What It Is, What It Means and Its Policy Uses* (pp. 135–150). Waterloo: Centre for International Governance Innovation.

Birdsall, N. (2016). Middle-Class Heroes: The Best Guarantee of Good Governance. *Foreign Affairs, 95*(2), 25–32.

Bratton, M. (2013a). Governance: Disaggregating Concept and Measurement. *APSA-Comparative Politics Newsletter, 23*(*1*), 12–14. Retrieved from http://comparativenewsletter.com/files/archived_newsletters/2013_winter.pdf.

Bratton, M. (2013b). *Measuring Government Performance in Public Opinion Surveys in Africa: Towards Experiments?* (United Nations University WIDER Working Paper 23/2013). Retrieved from https://www.wider.unu.edu/publications/working-papers/2013/en_GB/wp2013-023/.

Freedom House. (2013). *Freedom in the World 2013: Democratic Breakthroughs in the Balance. Selected Data from Freedom House's Annual Survey of Political Rights and Civil Liberties*. Washington, DC: Freedom House. Retrieved from https://www.freedomhouse.org/sites/default/files/FIW%202013%20Booklet.pdf.

Fukuyama, F. (2013a). What Is Governance? *Governance, 26*(3), 347–368.

Fukuyama, F. (2013b). *What Is Governance?* (Center for Global Development Working Paper 314). Washington, DC: Center for Global Development. Retrieved from https://www.cgdev.org/publication/what-governance-working-paper-314.

Gisselquist, R. M. (2012). *Good Governance as a Concept, and Why This Matters for Development Policy* (United Nations University WIDER Working Paper 30/2012). Retrieved from https://www.wider.unu.edu/publications/working-papers/2012/en_GB/wp2012-030/.

Gisselquist, R. M. (2015). Evaluating Governance Indexes: Key Criteria. In R. Rotberg (Ed.), *On Governance: What It Is, What It Measures, and Its Policy Uses* (pp. 23–54). Waterloo: CIGI Press.

Heller, N. (2013). How to Measure a Post-2015 MDG on Good Governance. *Global Integrity*. Retrieved from https://www.globalintegrity.org/posts/how-to-measure-a-post-2015-mdg-on-good-governance/.

Huntington, S. P. (2006). *Political Order in Changing Societies*. New Haven: Yale University Press.

Jerven, M. (2013). *Poor Numbers: How We Are Misled by African Development Statistics and What to Do about It*. Ithaca: Cornell University Press.

Kaufmann, D., Kraay, A., & Mastruzzi, M. (2010). *The Worldwide Governance Indicators: Methodology and Analytical Issues*. Washington, DC: Brookings Institute. Retrieved from http://www.brookings.edu/research/reports/2010/09/wgi-kaufmann.

Rotberg, R. I. (2004a). Strengthening Governance: Ranking Countries Would Help. *Washington Quarterly, 28*(1), 71–81.

Rotberg, R. I. (2004b). Improving Governance in the World: Creating a Measuring and Ranking System. In R. I. Rotberg & D. L. West (Eds.), *The Good Governance Problem: Doing Something about It* (pp. 3–30). Cambridge, MA: World Peace Foundation.

Rotberg, R. I. (2007). On Improving Nation-State Governance. *Daedalus, 136*(1), 151–155.

Rotberg, R. I. (2012). *Transformative Political Leadership: Making a Difference in the Developing World*. Chicago/London: The University of Chicago Press.

Rotberg, R. I. (2013). Governance Trumps Democracy: Examining the African Experience. In M. Boss, J. Moller, & S. E. Skaaning (Eds.), *Developing Democracies: Democracy, Democratization, and Development* (pp. 98–115). Aarhus: Aarhus University Press.

Rotberg, R. I. (2014). Good Governance Means Performance and Results. *Governance, 27*(3), 511–518.

Rotberg, R. I. (2015). The Governance of Nations: Definitions and Measures. In R. I. Rotberg (Ed.), *On Governance: What It Is, What It Measures, and Its Policy Uses* (pp. 7–22). Waterloo: CIGI Press.

Rotberg, R. I. (2017). *The Corruption Cure: How Citizens and Leaders can Combat Graft*. Princeton: Princeton University Press.

Rotberg, R. I., & Bhushan, A. (2015). The Indexes of Governance. In R. I. Rotberg (Ed.), *On Governance, On Governance: What It Is, What It Measures, and Its Policy Uses* (pp. 55–90). Waterloo: CIGI Press.

Rotberg, R. I., & Gisselquist, R. M. (2007). *Strengthening African Governance: The Index of African Governance*. Cambridge, MA: World Peace Foundation.

Rotberg, R. I., & Gisselquist, R. M. (2008). *Strengthening African Governance: The Index of African Governance*. Cambridge, MA: World Peace Foundation.

Rotberg, R. I., & Gisselquist, R. M. (2009). *Strengthening African Governance: The Index of African Governance*. Cambridge, MA: World Peace Foundation.

Rothstein, B. (2011). *The Quality of Government: Corruption, Social Trust, and Inequality in International Perspective*. Chicago: University of Chicago Press.

Rothstein, B., & Teorell, J. (2013). The Quality of Government: What It Is? How to Measure It? *APSA-Comparative Politics Newsletter, 23*(1), 1–5. Retrieved from http://comparativenewsletter.com/files/archived_newsletters/2013_winter.pdf.

Thomas, M. A. (2010). What Do the Worldwide Governance Indicators Measure? *European Journal of Development Research, 22*, 31–54.

United Nations Development Programme. (2000). *Strategy Note on Governance for Human Development*. New York: United Nations Development Programme. Retrieved from http://www.undp.org/content/dam/2007.

3

How International Rankings Constitute and Limit Our Understanding of Global Governance Challenges: The Case of Corruption

Alexander Cooley

Introduction

On December 2014, China's Ministry of Foreign Affairs publicly criticised Berlin-based anti-corruption NGO Transparency International (TI), following the release of its annual Corruption Perceptions Index (CPI), widely acknowledged as a leading global anti-corruption survey. The Chinese authorities were furious that in the watchdog's influential latest annual global rankings (the CPI), China had moved down 20 places, from 80 to 100, despite the fact that the authorities in Beijing were in the midst of a high-profile anti-corruption campaign aimed at taking down leading officials and targeting their activities and assets overseas. At the press conference, the Chinese Foreign Ministry spokesperson admonished that "[a]s a fairly influential international organisation, Transparency International should seriously examine the objectiveness and impartiality of its index" (Ministry of Foreign Affairs of the People's Republic of China, 2014).

China's public confrontation with Transparency International is an important example of the growing influence of international rankings and ratings in global governance. As this volume and others detail, over the last 20 years, dozens of new indexes have emerged that judge and compare states on different aspects of their governance (Cooley and Snyder 2015; Kelley and Simmons 2015). Proponents of rankings have viewed them as a way of highlighting

A. Cooley (✉)
Barnard College, Columbia University, New York, NY, USA

D.V. Malito et al. (eds.), *The Palgrave Handbook of Indicators in Global Governance*,
https://doi.org/10.1007/978-3-319-62707-6_3

state performance and spurring competition so as to improve governance. However, scholars have levelled a number of important critiques at the "rankings craze": rankings rely too heavily on sensational moralising and spotlighting rather than constructive diagnosis (Cooley and Snyder 2015); they privilege the technocratic and bureaucratic capacities and power of the North over its developing state counterparts (Davis et. al. 2012, 2015); they simplify complex social processes and project unwarranted normative assumptions; they imbue the authority issuing the ranking with unjustified (and unelected) power to rank and judge (Merry 2011); and they are methodologically fraught, often repackaging data of dubious quality in unwarranted and tenuous ways (Bhuta 2015; Thomas 2010).

Drawing on the above anecdote, this chapter adds yet another critique to this growing list: it criticises the methodological nationalisation that rankings confer on phenomena that are clearly transnational and global in scope. The approach of analysing corruption in terms of isolated country units has generated an incomplete and distorted policy picture of relatively clean developed states and corruption-prone developing states. The focus of the CPI and other governance indicators on the act, or rather the home jurisdiction, of bribe-taking overlooks the transnational networks that facilitate acts of grand corruption and kleptocracy in today's globalised world. Collectively, transnational channels that include shell companies, foreign real estate, and second citizenship and investor visa programmes serve to turn initially ill-gotten gains into ostensibly legitimate assets and legal rights. The Chinese case itself illustrates how corruption is instantiated in transnational networks of intermediaries that link actors, institutions, and processes across developing and developed states, and also shows how the Chinese authorities are transnationalising their sovereign authority and prosecutorial responses.

Corruption Rankings and the Distorting Lens of Methodological Nationalism

Corruption came onto the international policy agenda in the early 1990s in the wake of the end of the Cold War (Eigen 1996; Rose-Ackerman 1999; Wang and Rosenau 2001; Rothstein 2011; Vogl 2012; Gutterman 2014). At the national level, corruption had long been criminalised (in the United States as far back as the constitution), and thus neither corruption nor domestic legal and policy responses to it as such were new. What was new in the 1990s was seeing corruption as a ubiquitous problem among all states that required

an international normative response, rather than just a domestic policy problem amenable to local criminal investigative and regulatory solutions. Multilateral lending institutions, like the World Bank, donor governments, regional international clubs like the Organisation of American States, and a burgeoning range of transnational NGOs, all agreed that corruption was an international ill that required some sort of coordinated response within states, such as the establishment of anti-corruption agencies and the adoption of stronger anti-bribery laws. Overall, the conventional policy wisdom changed from seeing corruption as something that might be tolerated or even encouraged to facilitate international transactions to viewing it as a major barrier to economic development (Bukovansky 2006, 2015).

This shift to measuring and combatting corruption at the national level has been reinforced by the emergence, in both scholarship and policy-making circles, of international corruption and governance rankings. These include the previously mentioned Corruption Perceptions Index, compiled by Berlin-based NGO Transparency International, the World Bank's Worldwide Governance Indicators, and indices designed by various for-profit concerns (e.g. the Economist Intelligence Unit) (Bukovansky 2015; Merry et al. 2015; Andersson and Heywood 2009). These and other measures have created a view of corruption as primarily an act of domestic bribe-taking by public officials at the level of the nation-state.

State rankings, such as the CPI and the World Bank Governance Indicators, in turn, have been used by scholars as data sets to explore broad cross-national relationships involving corruption, as political scientists and economists in particular have disproportionately concentrated on large-*N* statistical approaches (Treisman 2007; Lambsdorff 2007; Rose-Ackerman 2010; Wedel 2012). The usual aim of such studies is to find what effect varying levels of (perceived) corruption have on some dependent variable of interest, most often economic growth, but also a whole range of other outcomes from democracy to infant mortality. A corollary is using these national corruption measures as the dependent variable with an eye to determining the causes of corruption (religious or colonial heritage, region, resource endowments, etc.) and how it can be best addressed (e.g. with a free press, development aid, a dedicated national anti-corruption agency, etc.). As a result, recurrent correlations in cross-national studies have entrenched a picture of relatively corrupt developing countries and relatively clean developed countries. However, importantly, even process-tracing single-case-study approaches generally share a methodological nationalism: corruption is conceived and measured in discrete bordered national units.

Therefore, while this scholarship has produced many insights, it has also tended to obscure the transnational nature of much large-scale corruption. Some important early policy responses to corruption were explicitly transnational. Perhaps the most significant of these was the OECD's anti-bribery convention (Abbott and Snidal 2002; Gutterman 2014). While bribing local officials had long been prohibited, bribing foreign officials was legal almost everywhere bar the United States up until at least the late 1990s. Yet even here the policy has in practice remained national. The signatory countries to the convention are reviewed for compliance on an individual and internal basis, and there is little or no effort by these states to coordinate with the states whose officials are bribed by OECD-based corporations (Stolen Asset Recovery Initiative, StAR 2014).

Interestingly, most news accounts of acts of bribery or corruption scandals uncovered in developing countries often mention the host country's rank in the CPI, especially if it is relatively low, but, unlike scholars, such news items usually also mention the nationality of an investing company; describe the investment vehicles, intermediaries, or legal schemes that facilitated the deal; and even trace the ultimate destination of the funds. Yet, our scholarly and policy lenses seem to de-emphasise these same transnational elements that straddle borders, despite the wealth of research on the topic (Stephenson 2016).

A Transnational Perspective on Corruption and Kleptocracy

Rather than being limited to one country, a sizeable proportion of large-scale corruption is conducted via transnational networks. In particular, large-scale corruption is almost always intermediated, rather than being simply an exchange between two parties. Critically, the intermediaries most commonly involved are legitimate businesses and actors providing for-fee services (bankers, lawyers, realtors), not criminal middlemen.

Although the transnational problem of Western multinational corporations (MNCs) bribing foreign government officials certainly remains important, this chapter outlines the other parts of the network that allow the laundering of such proceeds. Major corruption perpetrated by state leaders, senior officials, and their families is not just limited to receiving bribes from MNCs, and such kleptocracy almost always has an international aspect, as the criminal proceeds are moved abroad or in some cases never enter the klepto-

crat's home country at all (StAR 2011; Chayes 2015). Such cross-border transfers complicate the detection, investigation, and return of looted assets (Shelley 2014), and because the destinations for this money are typically large financial centres, like the United States, the UK, and Switzerland, these transfers are mediated by a wide range of professional intermediaries and provide options for conspicuous consumption. In the words of Alessandra Fontana of the U4 anti-corruption watchdog, "[C]orruption will remain a profitable crime in developing countries as long as counterparts in rich countries are willing to hide stolen resources" (Fontana 2011, p. i).

The various dimensions of a transnational network may include the following actors and activities:

- *Bribe-taker or kleptocrat*: a public official who engages in a corrupt act, using public resources or authority for private gain.
- *Bribe offeror*: a MNC, state, or international actor that offers a bribe to an official or seeks a contract or other form of preferential treatment.
- *Shell company*: a legal vehicle, the only purpose of which is to facilitate another transaction, usually camouflaging the identity of the actual transactors.
- *Brokers:* third-party intermediaries or professionals who serve to facilitate movements of funds from the original illegal act and officials along these transnational chains.
- *Destination banks*: the destination financial institutions in which corrupt proceeds eventually find themselves.
- *Citizenship providers*: the actors or institutions through which corrupt officials obtain the opportunity to evade the reach of their home jurisdiction and/or reside overseas.
- *Real estate and luxury assets*: properties or other luxury consumption goods purchased or invested in, usually through a shell company, designed to store and preserve the value of corrupt proceeds.

Of all these various nodes, corruption and governance indices overwhelmingly focus only on the first category. Perhaps remarkably, even the national origin of MNCs implicated in overseas corruption scandals is not accounted for in most international rankings of "clean" and "corrupt" countries, while the shell companies and professional intermediaries who facilitate these transactions are also disregarded. For example, New Zealand, which often tops the CPI as one of the world's cleanest countries, is also a global source of nominee directors of shell companies and trusts.

Why Are Professional Intermediaries Important?

The methodological nationalisation of corruption rankings tends to simplify a broad range of corruption crimes (passive and active bribery, extortion, embezzlement, nepotism, trading in influence) into a transaction involving a bribe-giver and recipient (Johnston 2005; Wedel 2012). In practice, however, major corruption offences typically involve a range of intermediaries. Transnational intermediaries and brokers have three main functions in the transnational corruption network. First, kleptocrats contract out the management of their stolen wealth (to "buy" rather than "make" in house). In doing so, they benefit from the specialised skills of professional intermediaries, while the fact that these intermediaries are themselves legal rather than criminal businesses reduces the danger of this opportunism. Second, these intermediaries serve to shield individuals engaged in corruption networks from accountability, as per the Global Wealth Chains thesis (Seabrooke and Wigan 2014). Whether they are bankers, accountants, residency consultants, or realtors, brokers serve to shield identities, create legal proxies for clients, obfuscate roles, and provide new spaces for corrupt officials to keep their wealth and to physically reside. There is an almost paradoxical dynamic here: as international state rankings and performance indicators have grown in popularity, an entire range of global industries has emerged enabling intermediaries to develop professional expertise and best practices to shield these same individuals from accountability and association with corruption in their home states. These global services and practices constitute the final link in the transnational corruption network: that of "exit" in the formulation of Albert Hirschman (1970). Breaking with the methodological nationalisation of corruption rankings not only necessitates looking at the transnational networks of bribe-givers, facilitating vehicles and brokers, but also the processes and methods through which corrupt officials manage to evade national regulators and law enforcement, launder money, and maintain legal protections if their criminal activities are detected by their home country authorities. The very right to flee and obtain the right to reside overseas, along with their ill-gotten assets, is itself a critical part of the grand corruption chain.

The Transnational Channels of Corruption

Transnational channels blur the distinctions between compartmentalised highly and poorly ranked economies and the spectrum of legal and illegal funds. Not all or even most of the foreign money flowing into North American

and Western European real estate represents the proceeds of criminal activity, yet it is absolutely true that international corruption rankings do not measure, capture, or reflect these critical processes. International banks serve both as ultimate destinations for ill-gotten gains and also routinely conduct multiple transactions with intermediaries and shell companies that facilitate the broad capital flight usually associated with highly corrupt jurisdictions. Mostly based on Swiss financial transaction data, Zucman (2015) estimates at least $7.6 trillion in global hidden wealth, with banks in Switzerland, Luxembourg, and the British Virgin Islands channelling and recycling most of these unaccounted funds. Thus, there is good reason to think that huge sums of criminal money continue to flow into financial centres like New York, London, Zurich, Singapore, and Dubai. Although the efforts of international banks to screen out dirty money have been drawing more watchdog attention (Global Witness 2009; US Senate 2010; Financial Services Authority 2011), this section will consider the equally important role of shell companies, real estate, and second citizenship providers.

Shell Companies

The leak in April 2016 of the documents detailing the clients of over 210,000 offshore companies established by the Panamanian-registered firm Mossack Fonseca—the so-called Panama Papers—dramatically spotlighted the problem of anonymous shell companies and Western complicity in shielding global assets. According to the International Consortium of Investigative Journalists, investigations leading from the 11.5 million leaked documents revealed the secret offshore holdings of 12 world leaders and at least 128 politicians (International Consortium of Investigative Journalists 2016).

Even prior to the Panama Papers, law enforcement officials and regulators had raised alarms over the central role of shell companies in facilitating corruption. For example, Assistant US Attorney General Lanny A. Breuer confirmed that “Shells are the No. 1 vehicle for laundering illicit money and criminal proceeds” (Wayne 2012). A StAR report (2011, p. 117–122) showed that the most common locations for the accompanying corporate bank accounts were the United States (107), Switzerland (76), and the UK (19). Of the 91 cases where there was sufficient information available, 72 involved a professional intermediary setting up a company. These data on the prime importance of major financial centres in hosting tainted wealth are further supported by an extensive field experiment undertaken by Findley et al. (2014), who, on the basis of over 7400 solicitations to company providers

around the world, found that OECD providers of shell companies were more likely than developing world tax havens to offer to sell shell companies without obtaining proper identifying documentation from the buyer. The authors further found that these companies tended to be created and marketed through chains of wholesale and retail corporate service providers.

Foreign Real Estate

Although the data are (as always) incomplete and subject to bias, a narrower focus on the purchase of international property strongly suggests that cities such as New York, Miami, London, and Paris are major destinations for a combination of licit and illicit wealth from the developing world. The transfer of this wealth is once again facilitated by networks of professional intermediaries, specifically real estate brokers and lawyers, who either are not required to identify the real owners of these purchases or regularly fail to do so in practice. Indeed, the whole process of luxury real estate purchasing employs a chain of individuals who are shielded from disclosure laws.

According to an inaugural luxury real estate survey released by Christie's in 2013, six North American cities (New York, Miami, San Francisco, Toronto, Los Angeles, and Dallas) ranked in the top ten of global luxury real estate markets (defined in terms of total annual listings over $1 million), with London, the Cote D'Azur, and Paris ranking first, second, and fifth, respectively. The only non-Western city was Hong Kong, ranking third, while Miami, London, and the Cote D'Azur topped the list of most international buyers in terms of overall percentage of sales. The 2014 annual report of the US National Realtors Association comments on how the types of home purchased by international buyers of US property typically differ from domestic ones, observing that international buyers acquire US property less in order to meet their "essential living needs" and more to establish an "individual's presence and standing in the community" (p. 12). The same report found that during the period April 2013–March 2014, international purchasers accounted for about $104 billion, or about 8 per cent of the total market, with international purchasers paying almost double both the mean and the median purchase prices of domestic buyers (pp. 5–6).

Industry analyses and reports by investigative journals make it clear that luxury real estate worldwide is increasingly being bought via shell companies using specialised real estate and law firms (Silverstein 2013; Rice 2014; Saul 2015). In each of these instances, a typical transaction might involve a foreign bank (often not in the official's home country) making a wire transfer to the

trust account of a law firm and/or real estate agent in the host country, who then buys the property in the name of a local or foreign shell company. Thus, a corrupt foreign official is distanced both from the source of the funds used to buy the property and the ownership of the property itself.

In the United States, lawyers and realtors do not have a duty to either establish the legitimacy of clients' funds or to report suspicious transactions to the authorities. After the passage of the Patriot Act in 2001, the US National Realtors Association fought for, and successfully secured, an exemption from a requirement that would have mandated brokers to conduct due diligence on purchasers (Saul 2015). Without a requirement of having to know the beneficial owners, real estate transactions via shell companies can effectively bypass bank reporting requirements.

One striking investigation into the buyers of condos at the Time Warner Center (Story and Saul 2015), one of Manhattan's most exclusive properties, found a "growing proportion of wealthy foreigners" in the building and identified 16 of them as having been subjects of government inquiries around the world, including four arrests and four fined for illegal activities. The report found that, while in 2003 about one-third of all purchases in the luxury building were made by shell companies, by 2014 the figure had increased to 80 per cent. Overall, since 2005 across the United States, the number of purchases of residences worth more than $5 million has tripled, totalling about $8 billion, while half of all of these luxury sales in 2014 were made to shell companies. Between April 2014 and March 2015, overseas buyers bought more than $100 billion worth of real estate in the United States (Frank 2015), and 55 per cent of these transactions were all cash.

The London market is even more penetrated by foreign and anonymous buyers than its New York and Miami counterparts. According to one report by Transparency International UK (2015a) based on police data, at least £122 billion of property in England and Wales is owned by companies with offshore registration. In London, of the 40,725 properties owned by foreign companies, 89 per cent were registered in tax havens, including the British Virgin Islands, Jersey, Guernsey, Luxembourg, and Panama. British Virgin Islands-registered entities accounted for about 5000 buyers from 2012 to 2014. Based on data collected from the UK property registry, the same report estimated that in Central London about 9.3 per cent of all properties in the City of Westminster were owned by offshore shell companies. An investigation in 2015 found 740 uninhabited properties, or "ghost mansions", in London that were worth £5 million or more each ("London's 3 bn Ghost Mansions", *Evening Standard*, February 14, 2014). In the UK, while lawyers and real estate agents do have a legal duty to scrutinise the seller, they hold no

such obligation towards the purchasing entity, regardless of whether it is a named individual or company registered offshore with no identified beneficial owner (Global Witness 2015).

Investor Visas and Second Citizenship Programmes

Complementing and assisting the spread of wealth across borders through networks of professional intermediaries and shell companies is the increasing importance of multiple citizenships or residencies, which give rise to the phenomenon of the globalised individual. Corrupt officials looking to exercise their exit options require a haven destination. Once more, intermediaries are important. Governments have long competed for mobile capital by wooing corporate investors, a dynamic often cast as the central leitmotif of the political economy of globalisation (Drezner 2008), but much less appreciated is the growing competition for mobile capital held by private individuals through investor visa, residency, and second citizenship programmes. For the states offering such facilities, the potential rewards are not only an infusion of foreign capital into the local economy but also a direct revenue gain for the recipient government. Such programmes have been overwhelmingly taken up by investors from China and the former Soviet states, and to a lesser extent by ones from elsewhere in the developing world. Second citizenships offer a level of security of person and property unavailable at home, either because the government is corrupt, repressive, and predatory or because the individual is a fugitive from justice, or both.

At the extreme, second (or subsequent) citizenship may involve a relatively simple exchange of money for a passport. One of the pioneers of this route, the Caribbean island federation of St Kitts and Nevis, set up its "economic citizenship" programme shortly after independence in 1984 (Abrahamian 2012). The passport entitles the holder to visa-free travel privileges to the European Union, Canada, and over 100 other countries, and applicants must either donate $250,000 to a government economic diversification fund or invest $400,000 in approved real estate that must be held for at least 5 years (most choose the latter; these totals do not include intermediaries' fees). There is no requirement to spend any time actually living in St Kitts and Nevis.

A common refrain is that such cash-for-passport deals are the exclusive province of tax havens and broken third-world countries, but the recent conduct of developed countries belies such sweeping stereotypes. Developed countries are keen to draw a sharp line between others selling citizenship and

their own investor visa programmes, but the difference is increasingly one of degree. EU members Cyprus, Malta, Bulgaria, Portugal, Spain, and others have rolled out new "golden residence permits" and "golden passport programmes" in the wake of the financial crisis (Shachar and Bauböck 2014, pp. 3-4).

Short of selling passports, a number of OECD countries have introduced new investor visa and residency programmes to allow high-net-worth investors safe residency in exchange for investments in the host country that exceed a certain threshold. In July 2015, Australia augmented its Significant Investor Visa scheme ($A5 million in approved investments followed by a four-year wait for permanent residency), with a Premium Investor Visa ($A15 million for permanent residency in a single year). Over 90 per cent of the first of these visas have been taken up by Chinese nationals. In the UK, the gateway investment is £2 million in approved investments, which entitles the holder to a 3-year stay, with a further £5 million then opening the path to permanent settlement. During the initial phase of the programme (2009–2013), a total of 1647 visas were granted (Migration Advisory Committee 2014, 22). The two leading recipient nationalities, together constituting about half of all permits, were Russian (433) and Chinese (419), with Egypt (46), India (44), and Kazakhstan (41), ranking fourth, fifth, and sixth, respectively. About half of all OECD countries have enacted some form of visa, residency, or citizenship for purchase or for investment. Law firms have also been involved in these trends, with certain firms openly advertising their "global citizenship" services, including marketing British and US citizenship as well as other EU member state and Caribbean options.

Corruption and Transnational Networks: China Revisited

To return to the chapter's opening anecdote, consider some vignettes about the extent to which corrupt Chinese officials fleeing the mainland have implicated themselves in these transnational channels and how the Chinese authorities have responded.

A People's Bank of China report written in 2008, and accidently leaked in 2011, revealed that in the period 1993–2008 between 16,000 and 18,000 Chinese officials had fled the country with $123 billion in state assets, with the leading destinations for senior cadres being the United States, Canada, and Australia (Arredy 2011). The report found that the greater the level of corruption of an accused official, the more likely they were to end up in a state

in the West that did not have an extradition agreement with China. In April 2015, China's Central Commission for Discipline Inspection (CCDI) publicly released a list of the 100 most wanted economic fugitives accused of economic crimes, including money laundering, bribery, and embezzlement, of which 66 were believed to be residing in the United States and Canada (Yan 2015).

Recently, Chinese nationals have been increasingly taking up the lion's share of major second citizenship and investor options. By the second half of 2015, Chinese nationals accounted for 37 per cent, or 1126, of the UK's Tier 1 Investor Visa programme (Transparency International UK 2015b, 11). In the United States in 2014, the large demand for EB-5 investor visas from Chinese nationals overran the programme's quota for the first time ever in its 24-year history: over 85 per cent of EB-5 funds in 2014 came from China (Shyong 2014). In terms of countries purchasing US real estate in 2014, buyers from China, Taiwan, and Hong Kong accounted for $28.6 billion of sales, by far the leading foreign buyers in the country (Searcey and Bradsher 2015).

The Chinese State Strikes Back Transnationally

Since his accession to power in December 2012, Xi Jinping has declared corruption among party/state officials a mortal threat to the regime's continued rule, and a campaign to root out such malfeasance has been given a correspondingly high priority. Much of this corruption has taken the form of officials moving their stolen assets, their families, and themselves to foreign countries and thus (they hope) to safety. Once again, foreign banks, shell companies, real estate, and second citizenships and residency permits are the key mechanisms—and none of these would have been detected or measured in the CPI or governance rankings.

Beijing has named one of its transnational initiatives, an effort since mid-2014 to seek the return of fugitive officials abroad, "Operation Fox Hunt". By its first anniversary, Fox Hunt had returned over 680 officials to China from 56 different countries (Gan 2015). Yet the programme was far from solving the basic problem of transnational corruption among Chinese officials, given that the people and assets returned are only a small fraction of those who fled. From April 2015, Beijing sought to re-double its efforts by subsuming Fox Hunt within a more encompassing effort known as "Sky Net". Together with the CCDI, Sky Net is an inter-agency effort involving the People's Bank of China (which houses the anti-money laundering agency), the Ministry of

Public Security, and the prosecutor's office (Fullerton 2015). Chinese embassies presented a "most wanted" list to the major host countries, with these countries' diplomatic representatives in Beijing also coming under pressure. Chinese police have entered the United States, Canada, Australia, and perhaps other states as tourists and then confronted the suspects directly, a practice that has led to some diplomatic tension (Mazzetti and Levin 2015). Thus, this new response to transnational corruption conflicts with traditional rules of inter-state behaviour.

The Jianjun Qiao Case: The Transnational Chinese Grain King

Consider how transnational networks facilitated the exit and overseas residence of one of China's most wanted economic fugitives in the case of Jianjun Qiao. Qiao was reportedly third on the Chinese government's list of the 100 most wanted fugitives accused of economic crimes that was released in April 2015, charged by the Chinese authorities of stealing $112 million. He is also accused by the US Justice Department of embezzling Chinese state funds through a series of fraudulent transactions from a large state-operated grain house that he operated from 1998 to 2011 in China's Henan Province (Saul and Levin 2015). Along with his alleged wife, Shilan Zhao (they were later found to be divorced), Qiao successfully applied for an investor visa in the United States in 2008 and then fled there following his indictment by Chinese officials in 2012 (Kuhn 2015).

In March 2015, Zhao was arrested in her home in Seattle and she and her ex-husband, still at large, were charged with money laundering and with fraudulently obtaining an EB-5 immigrant-investor visa to enter the United States. The couple had allegedly falsely represented themselves as "married" on their visa application and testified that their funds had been obtained by legal means. In early 2012, the couple had allegedly laundered $2.2 million in funds through a Canadian bank account and had used a shell company to funnel some of the proceeds to purchase a house in Newcastle, a Seattle suburb (Bhatt 2015). Qiao also reportedly purchased a condominium and house in Vancouver with all-cash payments; both of these purchases later became implicated in a money laundering investigation by Canadian authorities (O'Brien 2015). Chinese government authorities reportedly collaborated with the US Internal Revenue Service Unit and the Immigration and Customs Enforcement units that worked on the case (Areddy 2015).

Conclusions

It may well be the case that China is among the 100 "most corrupt" countries in the world, but this designation, both constituted and perpetuated by international rankings, tells us very little about how various transnational forces facilitate the relevant transactions in a globalised world. Nor do international rankings explain that much of this transnational activity appears ordinary and legal precisely because it is structured by a number of different international intermediaries—real estate agents and brokers, law firms specialising in citizenship, bankers, and consultants. Worse still, the methodological nationalisation inherent in international corruption rankings tends to frame policy prescriptions primarily as a function of improving governance, transparency, and monitoring within developing states. Certainly, Transparency International and other anti-corruption watchdogs are aware of these issues; indeed, TI UK has been one of the pioneers advocating tighter monitoring of real estate transactions and the investor visa programme (Transparency UK 2015a, b). However, it is the Berlin-based CPI that continues to draw the lion's share of attention, and the international media, policy makers, and actors involved in conducting due diligence religiously use the CPI as a tool to make judgements about clean and dirty states. Nationalisation appeals precisely because it places boundaries on a transnational phenomenon that makes it more amenable to selective external interventions and the deployment of anti-corruption "tools". Yet such interventions can only address a small portion of the transnational corruption chain, given that the Western-dominated service sectors and value chains that include banking, company providers, second citizenship services, and luxury real estate brokers are viewed as outside the scope of anti-corruption campaigns.

As the editors and many other chapter contributors argue, international indicators and rankings do not just simply reflect social and political realities. They simplify social phenomena, level unwarranted normative judgements, and selectively diagnose complex problems. Ranking states as discrete units, even as a form of social pressure designed to spur improvements, may actually obscure as much as it reveals, blinding us to other important sites of potential intervention that might actually improve global governance.

References

Abbott, K., & Snidal, D. (2002). Values and Interests: International Legalization in the Fight against Corruption. *Journal of Legal Studies, 31*(1), 141–178.

Abrahamian, A. A. (2012, February 12). Passports…for a Price. *Reuters*. Retrieved from http://www.reuters.com/article/us-passport-idUSTRE81B05A20120212

Andersson, S., & Heywood, P. (2009). The Politics of Perception: Use and Abuse of Transparency International's Approach to Measuring Corruption. *Political Studies, 57*(4), 746–767.

Arredy, J. (2011, June 16). Report: Corrupt Chinese Officials Take $123 Billion Overseas [Blog Post]. *Wall Street Journal*. http://blogs.wsj.com/chinareal-time/2011/06/16/report-corrupt-chinese-officials-take-123-billion-overseas/

Arredy, J. (2015, March 18). U.S. Indicts Ex-Chinese Government Official on Money Laundering Charges. *Wall Street Journal*. Retrieved from https://www.wsj.com/articles/u-s-indicts-ex-chinese-government-official-on-money-laundering-charges-1426688270

Bhatt, S. (2015, March 17). Newcastle Woman Charged with Fraud in Immigrant Investor Visas. *Seattle Times*. Retrieved from http://www.seattletimes.com/business/real-estate/newcastle-woman-charged-with-immigrant-investor-fraud/.

Bhuta, N. (2015). Measuring Stateness, Ranking Political Orders: Indices of State Fragility and State Failure. In A. Cooley & J. Snyder (Eds.), *Ranking the World: Grading States as a Tool of Global Governance* (pp. 85–111). Cambridge: Cambridge University Press.

Bukovansky, M. (2006). The Hollowness of Anti-Corruption Discourse. *Review of International Political Economy, 13*(2), 181–209.

Bukovansky, M. (2015). Corruption Rankings: Constructing and Contesting the Global Anti-Corruption Agenda. In A. Cooley & J. Snyder (Eds.), *Ranking the World: Grading States as a Tool of Global Governance* (pp. 60–84). Cambridge: Cambridge University Press.

Chayes, S. (2015). *Thieves of State: Why Corruption Threatens Global Security*. New York: W.W. Norton.

Cooley, A., & Snyder, J. (Eds.). (2015). *Ranking the World: Grading States as a Tool of Global Governance*. Cambridge: Cambridge University Press.

Davis, K. E., Kingsbury, B., & Merry, S. E. (2012). Indicators as a Technology of Global Governance. *Law & Society Review, 46*(1), 71–104.

Davis, K. E., Merry, S. E., & Kingsbury, B. (Eds.). (2015). *The Quiet Power of Indicators: Measuring Governance, Corruption, and Rule of Law*. New York: Cambridge University Press.

Drezner, D. (2008). *All Politics Is Global: Explaining International Regulatory Regimes*. Princeton: Princeton University Press.

Eigen, P. (1996). Combatting Corruption around the World. *Journal of Democracy, 7*(1), 158–168.

Financial Services Authority. (2011). *Banks' Management of High Money-Laundering Risk Situations*. London: The Financial Services Authority. Retrieved from http://www.fsa.gov.uk/pubs/other/aml_final_report.pdf.

Findley, M. G., Nielson, D. L., & Sharman, J. C. (2014). *Global Shell Games: Experiments in Transnational Relations, Crime and Terrorism*. Cambridge: Cambridge University Press.

Fontana, A. (2011). *Making Development Assistance Work at Home*. U4 Anti-Corruption Resource.

Frank, R. (2015, June 22). Wealthy Foreigners Bought $100 Billion in Real Estate. *CNBC*. Retrieved from http://www.cnbc.com/2015/06/22/wealthy-foreigners-bought-100-billion-in-us-real-estate.html

Fullerton, J. (2015, March 29). Operation Skynet: China's Anti-Corruption Campaign Goes International. *The Independent*. Retrieved from http://www.independent.co.uk/news/world/asia/operation-skynet-chinas-anti-corruption-campaign-goes-international-as-beijing-reaches-out-to-10142310.html

Gan, N. (2015, April 18). Revealed: The Team Behind China's Operation Fox Hunt. *South China Morning Post*. Retrieved from http://www.scmp.com/news/china/policies-politics/article/1768993/team-behind-chinas-operation-fox-hunt-revealed

Global Witness. (2009). *Undue Diligence: How Banks Do Business with Corrupt Regimes*. London: Global Witness. Retrieved from https://www.globalwitness.org/en/campaigns/corruption-and-money-laundering/banks/undue-diligence/.

Gutterman, E. (2014). The Legitimacy of Transnational NGOs: Lessons from the Experience of Transparency International in Germany and France. *Review of International Studies, 40*(2), 391–418.

Hirschman, A. O. (1970). *Exit, Voice and Loyalty: Responses to Decline in Firms, Organizations and States*. Cambridge, MA: Harvard University Press.

International Consortium of Investigative Journalists. (2016). *The Panama Papers*. Washington, DC: International Consortium of Investigative Journalists. Retrieved from https://panamapapers.icij.org/.

Johnston, M. (2005). *Syndromes of Corruption: Wealth, Power and Democracy*. Cambridge: Cambridge University Press.

Kelley, J. G., & Simmons, B. A. (2015). Politics by Number: Indicators as Social Pressure in International Relations. *American Journal of Political Science, 59*(1), 55–70.

Kuhn, A. (2015, June 24). *When Corrupt Chinese Officials Flee, The U.S. Is A Top Destination*. National Public Radio. Retrieved from http://www.npr.org/sections/parallels/2015/06/23/416828057/when-corrupt-chinese-officials-flee-the-u-s-is-a-top-destination

Lambsdorff, J. G. (2007). *The Institutional Economics of Corruption and Reform*. Cambridge: Cambridge University Press.

Mazzetti, M. & Levin, D. (2015, August 15). Obama Administration Warns Beijing about Covert Agents Operating in U.S. *New York Times*. Retrieved from https://www.nytimes.com/2015/08/17/us/politics/obama-administration-warns-beijing-about-agents-operating-in-us.html

Merry, S. E. (2011). Measuring the World. *Current Anthropology, 52*(S3), S83–S95.

Merry, S. E., Davis, K. E., & Kingsbury, B. (Eds.). (2015). *The Quiet Power of Indicators: Measuring Governance, Corruption, and Rule of Law*. Cambridge: Cambridge University Press.

Migration Advisory Committee. (2014). *Tier 1 (Investor) Route: Investment Thresholds and Economic Benefits*. London: Migration Advisory Committee.

Ministry of Foreign Affairs of the People's Republic of China. (2014, December 3). *Foreign Ministry Spokesperson Hua Chunying's Regular Press Conference* [Press Release]. Retrieved from http://www.fmprc.gov.cn/mfa_eng/xwfw_665399/s2510_665401/t1216399.shtml

O'Brien, F. (2015, April 7). Property Sales Spike Money Laundering Fear. *Business Vancouver*. Retrieved from https://www.biv.com/article/2015/4/property-sales-spike-sparks-money-laundering-fears/

Rice, A. (2014, June 29). Stash Pad. *New York Magazine*. Retrieved from http://nymag.com/news/features/foreigners-hiding-money-new-york-real-estate-2014-6/

Rose-Ackerman, S. (1999). *Corruption and Government: Causes, Consequences and Reform*. Cambridge: Cambridge University Press.

Rose-Ackerman, S. (2010). The Law and Economics of Bribery and Extortion. *Annual Review of Law and Social Science, 6*(1), 217–238.

Rothstein, B. (2011). Anti-Corruption: The Indirect "Big Bang" Approach. *Review of International Political Economy, 18*(2), 228–250.

Saul, S. (2015, March 10). Treasury Urged to Scrutinise Foreign Real Estate Buyers for Money-Laundering Risk. *New York Times*. Retrieved from https://www.nytimes.com/2015/03/11/business/treasury-urged-to-scrutinize-foreign-real-estate-buyers-for-money-laundering-risk.html

Saul, S., & Levin, D. (2015, May 15). Charged with Graft in China, Some Fugitives Are Finding Luxury in U.S. *New York Times*. Retrieved from https://www.nytimes.com/2015/05/16/world/asia/china-hunts-fugitives-accused-of-corruption-many-in-us.html

Seabrooke, L., & Wigan, D. (2014). Global Wealth Chains in the International Political Economy. *Review of International Political Economy, 21*(1), 257–263.

Searcey, D., & Bransher, K. (2015, November 28). Chinese Cash Floods U.S. Market. *New York Times*. Retrieved from https://www.nytimes.com/2015/11/29/business/international/chinese-cash-floods-us-real-estate-market.html

Shachar, A., & Bauböck, R. (2014). *Should Citizenship Be for Sale?* (European University Institute Working Paper RSCAS 2014/1). Florence: European University Institute.

Shelley, L. (2014). *Dirty Entanglements: Corruption, Crime, and Terrorism*. New York: Cambridge University Press.

Shyong, F. (2014, August 28). Visa Program for Wealthy Investors Maxed Out by Chinese Demand. *Los Angeles Times*. Retrieved from http://www.latimes.com/local/la-me-0830-chinese-visas-20140830-story.html

Silverstein, K. (2013, October 2). Miami: Where Luxury Real Estate Meets Dirty Money. *The Nation*. Retrieved from https://www.thenation.com/article/miami-where-luxury-real-estate-meets-dirty-money/

Stephenson, M. (2016). *Bibliography on Corruption and Anti-Corruption*. Retrieved from Harvard University, Law School: http://www.law.harvard.edu/faculty/mstephenson/2016PDFs/Stephenson%20Corruption%20Bibliography%20May%202016.pdf.

Stolen Asset Recovery Initiative (StAR). (2011). *The Puppet Masters: How the Corrupt Use Legal Structures to Hide Stolen Assets and What to Do about it*. Washington, DC: World Bank/United Nations Office on Drugs and Crime. Retrieved from https://star.worldbank.org/star/sites/star/files/puppetmastersv1.pdf.

Stolen Asset Recovery Initiative (StAR). (2014). *Left Out of the Bargain: Settlements in Foreign Bribery Cases and Implications for Asset Recovery*. Washington, DC: World Bank/United Nations Office on Drugs and Crime. Retrieved from http://star.worldbank.org/star/sites/star/files/9781464800863.pdf.

Story, L., & Saul, S. (2015, February 7). Stream of Foreign Wealth Flows to Elite New York Real Estate. *New York Times*. Retrieved from https://www.nytimes.com/2015/02/08/nyregion/stream-of-foreign-wealth-flows-to-time-warner-condos.html

Thomas, M. A. (2010). What Do the Worldwide Governance Indicators Measure? *European Journal of Development Research, 22*(1), 31–54.

Transparency International United Kingdom. (2015a). *Corruption on Your Doorstep: How Corrupt Capital Is Used to Buy Property in the UK*. London: Transparency International United Kingdom. Retrieved from http://www.transparency.org.uk/publications/corruption-on-your-doorstep/.

Transparency International United Kingdom. (2015b). *Gold Rush: Investment Visas and Capital Flows in the UK*. London: Transparency International United Kingdom. Retrieved from http://www.transparency.org.uk/publications/gold-rush-investment-visas-and-corrupt-capital-flows-into-the-uk/.

Treisman, D. (2007). What Have We Learned about the Causes of Corruption from Ten Years of Cross-National Research? *Annual Review of Political Science., 10*(1), 211–244.

U.S. Senate. (2010). *Keeping Foreign Corruption Out of the United States: Four Case Histories. Permanent Subcommittee on Investigations*. Washington, DC: U.S. Senate. Retrieved from http://www.hsgac.senate.gov.

Vogl, F. (2012). *Waging War on Corruption: Inside the Movement Fighting the Abuse of Power*. Lanham: Rowman and Littlefield.

Wang, H., & Rosenau, J. N. (2001). Transparency International and Corruption as an Issue of Global Governance. *Global Governance, 7*(1), 25–49.

Wayne, L. (2012, June 30). How Delaware Thrives as a Haven. *New York Times*. Retrieved from http://www.nytimes.com/2012/07/01/business/how-delaware-thrives-as-a-corporate-tax-haven.html

Wedel, J. R. (2012). Rethinking Corruption in an Age of Ambiguity. *Annual Review of Law and Social Science, 8*(1), 453–498.

Yan, S. (2015, April 23). China Releases "Most Wanted" List of Economic Fugitives. *CNN Money*. Retrieved from http://money.cnn.com/2015/04/23/news/china-most-wanted-economic-fugitives/index.html

Zucman, G. (2015). *The Hidden Wealth of Nations: The Scourge of Tax Havens*. Chicago: University of Chicago Press.

4

Measuring Governance As If Institutions Matter: A Proposal

José Antonio Alonso and Carlos Garcimartín

Introduction

During the past two decades, an ample literature coming from different analytical approaches has underlined the crucial role that institutions play in the process of economic and social change. "Institutions matter" has hence become a new mantra in development studies, and with it the study of the morphology and functions of the state and the conditions of governance has gained new interest. Accompanying this process, a considerable number of institutional quality indicators have been elaborated by multilateral institutions, risk-rating agencies, academic centres, and non-governmental organisations. Given the extent of the available repertoire, it is not surprising that the characteristics and quality levels differ greatly among indicators. As a common feature, however, most of them lack a theoretical framework linking the indicators to a previously well-defined analytical approach.

In the specialised literature, some of these new sets of indicators have been indifferently used to provide empirical tools for the analysis of governance, institutional quality, and the performance of the state. Although there are

J.A. Alonso (✉)
Universidad Complutense de Madrid, Madrid, Spain

C. Garcimartín
Universidad Rey Juan Carlos, Madrid, Spain

D.V. Malito et al. (eds.), *The Palgrave Handbook of Indicators in Global Governance*,
https://doi.org/10.1007/978-3-319-62707-6_4

some commonalities in these three fields, we should stress that they refer to different realities. As Greif argues (2006), *institutions* might be defined as a set of social factors, rules, beliefs, values, and organisations that jointly motivate regularity in individual and social behaviour. Institutions are, therefore, related to formal and informal factors that shape economic and social actors' behaviour. The *state* is a part of the institutional framework; probably the most relevant part, since it defines the explicit (or formal) rules under which the rest of the actors operate, has the capacity to enforce rules, and provides public goods and services to society. Finally, *governance* is the set of social structures and relations needed to pursue collective goals through diverse means of operation and coordination. It not only refers to action by the state, but it also includes the relations established among social actors—state, private sector, and civil society—to articulate responses to collective problems. In other words, governance refers to the way in which collective action is built in the public realm.

But how should "good governance" be measured? There is not an easy response to this question, among other reasons because governance is still a complex and hazy concept and we cannot measure what we have not yet adequately conceptualised. In this situation, a resort to empirical analysis might not be a solution. We could use the available databases, trying to identify some empirical regularities, but we cannot avoid the results being as questionable as the theoretical justification for the indicators employed. We definitely need more conceptual discussion of governance in order to advance with the task of measuring "good governance".

As suggested above, we identify governance as the collection of factors and relations that makes it possible to build social responses to collective problems. This definition is wider than that adopted by Fukuyama (2013). In a very well-known paper, Fukuyama (2013, p. 3) links governance with "government's ability to make and enforce rules, and to deliver services, regardless of whether that governance is democratic or not". As some critics have suggested, Fukuyama's vision is too narrow and more in tune with the concept of "governability" than with "governance", as the latter goes beyond government behaviour and embraces the relationship between government, the private sector, and civil society (see, among others, Levi-Faur 2013). Our definition is more in line with that offered by Kooiman, who characterises governance as "the totality of interactions in which public and private actors participate, aimed at solving societal problems or creating societal opportunities" (2003, p. 4).

Given this situation, in order to measure governance we adopt a prudent strategy which goes to the base of institutional analysis. In our view, given that

the state is part of the available institutional framework and governance is influenced by institutions, the criteria (properties) that, broadly speaking, define the quality of governance must be the foundations on which the evaluation of both, the state and governance, rest. In other words, we go back to the fundamentals that enable collective action, trying to understand good governance as the result of high-quality institutions shaping social interaction in the public realm.

Going beyond the available indicators, our main goal in this chapter is to provide a practical guide that allows us to develop indicators measuring the quality of institutions and, by doing so, to investigate governance. With this aim, we first present the criteria that, in our view, define the quality of institutions. Second, we summarise the findings of a previous empirical exercise developed by the authors and used to test the explanatory capacity of the criteria suggested. Third, we provide some ideas and examples that can be useful to transform our analytical approach into governance indicators. Finally, we present some policy implications drawn from the analysis.

Institutions and Development

Economists have traditionally tended to identify the causes of development in terms of resource endowment and technology. In essence, modern growth theory, built on the seminal contribution by Solow (1956), responds to this notion. Unlike this vision, a new perspective, not necessarily incompatible, has emerged in last decades. This perspective insists on the relevance of normative frameworks and institutions in fostering development (Acemoglu et al. 2005; North 1990, 2005; Fukuyama, 2011a, b, 2014). The institutional structure defines incentives and penalties, shapes social behaviour, and enables collective action, thus conditioning development. In recent years, a myriad of empirical studies have supported this relationship between institutional quality and development (see, e.g. Hall and Jones 1999; Acemoglu et al. 2002; Rodrik et al. 2002; Alonso 2011).

However, from an economic policy perspective it is not sufficient to acknowledge that institutions do matter. It is also necessary to identify the determinants of institutional quality. This is a crucial task for policies aimed at building better institutions to be implemented. Nevertheless, empirical research is scarcer in this area and its conclusions are less tenable. This is partly due to the difficulties faced by empirical work in this field. More precisely, it faces (i) deficient institutional quality indicators, (ii) problems stemming from endogenous variables, (iii) collinearity among the potential explanatory

variables preventing them from being considered independent factors, and (iv) the possible presence of omitted variables that can bias the parameters estimated.

A simplifying assumption about institutions made by economists in their modelling also hinders empirical research. Institutions are often considered to be efficient responses to transaction costs (Furubotn and Richter 2005). It is assumed that agents operate exclusively driven by rational optimisation criteria, that social dynamics eliminate inefficient institutions and that the existing ones improve social welfare. As a consequence, there is no problem whatsoever in defining "ideal institutions"—those of successful countries—and in transplanting them to other nations. These premises have inspired a good proportion of international donors' institutional reform programmes. At the same time, indicators based on this conception assume that institutions emerge as substitutes for markets when markets are highly inefficient (with high transaction costs). Accordingly, the quality of the former can be measured by the same criteria as the quality of the latter. Then, flexibility and limited intrusion into agents' behaviour emerge as the main criteria to evaluate institutional quality in several indicators.

This conception is encouraged by both the rational choice doctrine and the neo-institutionalist approach, the two dominant institutional conceptions in economics. In the first case, because it is assumed that institutions (and particularly states) are predatory, and therefore it is necessary to limit their discretionary action (Buchanan and Tullock 1962). In the second case, it is because it is assumed that efficient institutions can be built by manipulation of incentives, applying market-like procedures to the public sector (Moe 1990). As a consequence, these approaches are prone to identifying good governance with "the norms of limited government that protect private property from predation by State" (Kaufmann et al. 2007, p. 2). Such a narrow conception, however, does not seem to account either for some cases that have often been considered examples of good governance—such as Sweden and Demark—or for some others that have achieved visible success in their development processes—such as South Korea and China.

One of the problems that affect the aforementioned analyses is that they conceptualise governance exclusively (or mainly) in a principal-agent framework (as economists usually do). However, we would like to underline the view of governance as mechanisms for enabling collective action (Booth and Cammack 2013). In the first case, limits, checks, and controls on a (predatory) state are emphasised; in the second, institutions emerge as mechanisms for solving problems of social coordination, enabling collective action. Both

approaches are useful and complementary, but an approach balanced between them is needed. At least, as Scott suggests, "institutions work both to constrain and empower social behaviour" (2008, p. 52).

Another discrepancy relates to the different ways in which institutions are conceived, which affects the role that an institution should supposedly play. Basically, economists have understood institutions in two different ways: either as exogenous constraints on agents' behaviour or as endogenous self-enforcing rules (Greif and Kingston 2011). In both cases, institutions do not really work if they are not capable of shaping social behaviours, but the relation between rules and enforcement is very different in each case. In the first case (see North 1990), enforcement of the rules is considered a separate issue from the formation and content of the rules themselves. Rules respond to the interests and needs of their creators and enforcement relies on a third party, with it being costly to build institutional structures to ensure that rules are followed. In the second case (see Greif 2006; Aoki 2001), enforcement is mainly endogenised and institutions are conceived as equilibria in repeated games. Agents create the structure that provides the motivation to act in accordance with the behaviours predicted by institutions, perpetuating their structure. Therefore, in order to evaluate institutions it is important to analyse not only the rules designed, but also individuals' motivations to obey them (and the expected behaviour of others). In other words, it is as important to study the incentives framework in which agents operate as why they behave according to them, which obliges discussion of why some rules are observed while some others are not. Hence, the credibility and legitimacy of institutions becomes a basic feature conditioning their efficiency. In sum, other criteria (in addition to or different from flexibility) should be taken into account if institutional quality is to be measured.

Measuring Institutional Quality

The effects of institutional quality on the development process have been the subject of an increasing and active area of research. To make this possible, a considerable number of institutional quality indicators have been elaborated. We should be grateful for this effort even if the proliferation of indicators has not been balanced by a similar effort to clarify them theoretically. The relevance, quality, and accuracy of these indicators are very diverse, but, broadly speaking, we could define the situation as far from satisfactory (Arndt and Oman 2006; Alonso and Garcimartín 2008; Thomas 2009).

Among the main problems faced by most of the available indicators, six seem to be of particular importance:

- *Most of the available indicators are subjective.* Perception-based data are helpful in capturing information about topics that are difficult to grasp with objective measures, but they can be subject to frequent variations due to changes in public opinion with no relation to a real change in the structural phenomenon that they are supposed to capture. For example, corruption is frequently measured using the Corruption Perceptions Index (CPI), a subjective indicator created by Transparency International that is a component of several indexes and composite indicators on governance. Based on this indicator, Italy climbed in the corruption ranking (showing higher index results than Colombia and Mexico) around 1995, basically as a consequence of the effect that some important trials against corruption had on public opinion (in the framework of the *Tangentopoli* judicial campaign). Something similar is currently happening in Spain as a consequence of the effect of well-publicised judicial operations against some politicians involved in corruption. In these cases, the CPI reflects changes in public opinion rather than actual changes, with the role played by the media in this respect being crucial. In addition, this type of indicator can be biased due to an inadequate sample construction. For example, the quality of the institutional framework of the Global Competitiveness Index, launched by the World Economic Forum, is based on a survey of top business executives in all the countries considered. Beyond the relevance of these opinions, we must admit that they are far from representative of all the social sectors affected by the institutional framework. In other cases, the index is based on the subjective opinions of a group of experts (whose representativeness is debatable). In these cases, since the concept of "institutional quality" is not well defined, different experts may understand different things when responding to the same survey question (Fukuyama 2013).
- *Ideological bias.* Some indicators are based on assumptions that are highly questionable. The Economic Freedom Index, launched by the Heritage Foundation, is probably the clearest example. Any type of state interference or regulation is penalised in this index. As a consequence, it is no wonder that France is ranked 73, far behind countries such as Kazakhstan, Colombia, Peru, Mexico, Bahrain, and Malaysia. In the case of the World Bank's Doing Business project, for example, the ranking in the tax-paying component is headed by Hong Kong and the United Arab Emirates, while, for example, Sweden is ranked 37, only because tax regulation on corporations is more demanding in the latter country. In these cases, market flexi-

bility is considered the main criterion for assessing institutional quality. However, institutions are created not only (not even mainly) to ensure market flexibility but also to reduce uncertainty and to guarantee a reasonable distribution of the results of collective actions (which could require a reduction in market flexibility).

- *Confusion between outcomes and criteria.* A good institution can produce bad outcomes and a bad institution can produce good outcomes (see Glaeser et al. 2004). The cases of Franco's regime in Spain and Pinochet's regime in Chile can be examples of bad institutional frameworks (not only for democratic reasons) that were able to produce good economic results. We should distinguish between the characteristics of an institution and the result of the policies that this institution promotes. In general terms and for long periods we should expect a positive correlation between institutions and outcomes, but the link is far from mechanical. Remembering what Popper said regarding democracy, the main difference between a good and a bad institution is not that the former is able to avoid mistakes, but that it is in a better condition to identify and correct its own mistakes.
- *Aggregation.* Governance and institutional quality represent a multifaceted reality, and therefore it is not possible to capture the entire reality of governance using one single indicator. Most indexes and composite indicators use several indicators of specific aspects of institutional quality as the basis for building an aggregate index. However, without clear inferences about the relationships among the variables considered, any criterion for aggregation will be highly debatable. Additionally, after aggregation, it can be hard to know what the indicator is actually measuring, given that it integrates variables related to many different dimensions (e.g. in the Worldwide Governance Index (WGI) we can simultaneously find outcomes, processes, and policy choices).
- *Low quality of information.* The information on which indicators are based is not always adequate or it experiences frequent changes over time. This usually happens with indicators based on polls, because of frequent changes in the samples. Since quality varies across countries, international comparisons are not always reliable.
- *Lack of a theoretical model.* Most indicators are built without any underlying theoretical model supporting their construction. The information collected is supposed to be linked to governance or institutional quality, but the theoretical framework to which this information should respond is not made explicit or it simply does not exist. For example, one of the most widely used databases, the WGI (as well as other sets of indicators), is in fact the combination of many indicators with many underlying theories and normative perspectives (Andrews 2008, p. 382).

Some of the aforementioned problems are aggravated by a visible impatience to give empirical support to a conceptual reality that is still blurred and under construction. It is unlikely that we can properly measure what we have not been able to conceptualise; at most we will have indicators, but without any theory explaining what we measure (Andrews 2008). A more dangerous problem derived from this lack of critical conceptualisation is the temptation to use these indexes to give way to isomorphism and support "monocropping", as if a universal single best institutional framework existed (Evans 2004).

To overcome these problems, we suggest returning to the theoretical criteria associated with good institutions and basing the search for appropriate indicators on a sound exploration of the fundamentals. That is, prior to building indicators we should try to answer the following key question: what do we demand from institutions?

Institutional Quality Criteria

As suggested above, institutions shape human behaviour and social interactions, and by doing so reduce uncertainty and enable collective action. According to this view, institutions have two basic economic functions: on the one hand to reduce transaction costs, granting certainty and predictability to social interactions; and on the other hand, to facilitate coordination of economic agents. If both functions are kept in mind, institutional quality can be defined in terms of four basic properties:

- *Capacity (static efficiency).* The institution's ability to promote equilibria that achieve the maximum social returns permitted by the technological frontier. Room for rent-seeking activities is an example of low static efficiency.
- *Credibility (or legitimacy).* The institution's capacity to define intertemporal credible contracts, that is to say the institution's ability to generate rules that are effectively internalised in agents' behaviour (Alonso and Garcimartín 2013, p. 209). This property of institutions refers to three complementary criteria:
  - *Legitimacy and fairness.* Institutions must create the perception that their actions are desirable and adequate within the framework of some social construction of norms, values, and beliefs (Suchman 1995, p. 574).

  - *Transparency and accountability*. Institutions must be transparent and accountable to stakeholders.
  - *Enforcement capacity*. In order to be credible, institutions must have the capacity to enforce rules, either because they apply incentives or penalties when necessary or because they represent self-enforcement equilibria.

- *Stability (or predictability)*. An institution fulfils its function if it reduces the uncertainty associated with human interaction. In fact, one of the institutional functions is to grant a higher level of safety and stability to social relations by diminishing uncertainty.
- *Adaptability (or dynamic efficiency)*. Social and economic reality is continuously changing. An institution must be able to anticipate social changes or at least to generate the incentives that facilitate agents' adjustment to these changes (North 2005).

In our view, institutional quality indicators should provide information about these four dimensions of institutions. And these same dimensions should be considered when good governance has to be assessed. However, it is worth mentioning that frictions or conflicts may exist among these dimensions, and certain trade-offs can emerge in the implementation of any institutional choice. For example, if stability and security are prioritised, this may result in lower dynamic efficiency, since this requires a capacity to adapt to future changes.

An Empirical Exercise

Before applying this approach to develop an analytical framework that can be used to measure governance, we will present a brief summary of the results of a prior empirical study conducted by the authors to test the explanatory capacity of the above four criteria as determinants of institutional quality (Alonso and Garcimartín 2013). However, we must clarify that this investigation was just a first step to test the potential of this approach, since no institutional quality indicator (or set of indicators) built on the basis of these four dimensions is currently available.

Our strategy followed three steps. First, we identified the most sound and reliable indicators to evaluate institutional quality using different alternatives (including the six dimensions of the World Bank's Governance Indicators). Second, we identified variables that served as proxies for the four criteria

mentioned above. Third, we assumed that if the empirical work supported the role of these variables as determinants of institutional quality, then it would provide preliminary support for our approach to defining institutional quality.

However, before developing our analysis, we studied the impact of some other determinants traditionally considered in the literature on institutional quality. Since institutions are inertial by nature, researchers have tended to investigate as potential determinants variables that are relatively constant over time, such as those related to the geographical or historical features of countries (see La Porta et al. 1999; Chong and Zanforlin 2000; Acemoglu et al. 2001; Islam and Montenegro 2002; Easterly and Levine 2003; Alesina et al. 2003; Easterly et al. 2006; Rigobon and Rodrik 2004). The main candidates have been the following:

- *Ethno-linguistic fragmentation.* It is assumed that greater (ethnic and linguistic) heterogeneity may fuel tensions and conflicts between different groups, reduce social cooperation, and generate a mismatch between formal and informal institutions.
- *Origin of the legal system.* It is argued that the British legal system (common law), and to a lesser extent the German and Scandinavian systems, is based on a greater recognition of economic freedom. On the contrary, the French legal system (civil law) and even more the Soviet system were designed to reinforce the state. Accordingly, British and Nordic legal traditions have been expected to be associated with higher institutional quality.
- *Identity of the coloniser.* It is assumed that UK colonisation has been less noxious than others, since it has favoured the emergence of an institutional framework better prepared to foster a market economy.
- *Geographical factors.* It is considered that the location of a country in the tropics, access to the sea, or soil fertility may have influenced the development of strong quality institutions.
- *Valuable natural resources*. Resources can negatively affect institutions by fostering rent-seeking activities and replacing tax revenues with other revenue sources which are less transparent and less subject to accountability.

We estimated the impact of these traditional variables on institutional quality and only one of them turned out to be significant: geographical location (proxied by latitude) (Table 4.1). Although our results could partially differ from the findings of other studies due to differences in the indicators and samples employed, we found that these variables lost significance once

Table 4.1 Determinants of institutional quality (I) (dependent variable: institutional quality)

Variable	Value	*t*- Ratio
Constant	−4.58	9.43
Per capita GDP	0.50	7.82
Ethnic fragmentation	0.01	1.12
Former British colony	0.18	1.63
Former Spanish colony	−0.22	1.65
Former French colony	0.05	0.33
British legal system	−0.09	0.86
Geographical location	1.43	3.52
Fuel	−0.11	0.64
Middle East & North Afr.	−0.71	4.00
Europe & Central Asia	−0.59	4.35
Adjusted R^2	0.77	
No. of observations	127	

Source: Alonso and Garcimartín (2013, p. 216)
Endogenous: per capita GDP 2004. Instruments: pc GDP 1990
Instrumental variables. Robust estimates. Exactly identified equation

the estimations are controlled for per capita income. In other words, the colonial origin or the legal system is only significant when the level of development is omitted from the estimation.

The alternative model estimated in the study considered potential determinants related to the previously defined institutional quality criteria: static efficiency, dynamic efficiency, predictability, and legitimacy. Five variables were considered:

- *Development level*. This is one of the first clear explanatory variables, and it influences institutional quality through both supply and demand factors. First, it determines the availability of resources to build good institutions. Second, it generates a larger demand for quality institutions. It is a determinant mainly related to the static efficiency of institutions (and possibly to dynamic efficiency too).
- *International openness*. This is related to the dynamic efficiency of institutions. First, it creates a more dynamic, sophisticated, and demanding environment, which fuels a greater demand for good institutions. Second, international openness encourages a more competitive environment; it can therefore hinder rent-seeking activities. Finally, openness can facilitate learning processes and imitation of good practices from the experience of other countries.
- *Income distribution*. This is claimed to affect both institutional predictability and legitimacy. First, because strong inequality causes divergent inter-

ests among different social groups, which in turn leads to conflicts, sociopolitical instability, and insecurity. Second, inequality facilitates the capture of institutions by power groups, whose actions are orientated to particular interests rather than to the common good. Third, it diminishes the disposition of social agents to act cooperatively and favours corruption and rent-seeking activities.

- *Education*. This variable is related to the dynamic efficiency of institutions. A better-educated population demands more transparent and dynamic institutions and permits them to be built (because human capital is an important input in the creation of high-quality institutions).
- *Taxation*. Although this variable had not been taken into account in previous studies, in our opinion it was a crucial variable that could affect both the static efficiency and the legitimacy of institutions. As the fiscal sociology literature suggests, "sources of State revenue have a major impact on patterns of State formation" (Moore 2004, p. 297). A sound tax system not only provides the necessary resources to build high-quality institutions, but also enables the consolidation of a social contract that gives rise to a more demanding relationship between state and citizens. As a result, there will be a higher degree of transparency and accountability, which leads to better institutional quality. This may not happen with public revenues collected from other sources such as state-owned companies or natural resources.

Our finding was that these variables explained a large proportion of institutional quality differences across countries. All of them showed the expected sign and were significant, except the openness rate. In particular, the results suggested that a higher level of development leads to greater institutional quality, but as the latter also promotes economic development this implies that the two variables interact and may lead to vicious or virtuous circles of institutional quality and growth. On the other hand, a more equitable income distribution improves institutional quality, and, on the contrary, a context of high social inequality leads to bad institutions. With regard to taxation, according to our findings a sound tax system promotes institutional quality, since it provides the necessary public revenue to build good institutions, and it creates a more direct and demanding relationship between citizens and the state. Finally, education seemed to have a positive impact on institutional quality (Table 4.2).

In sum, the empirical evidence suggested that the chosen variables related to the aforementioned criteria were good determinants of institutional quality.

Table 4.2 Determinants of institutional quality (II) (dependent variable: institutional quality)

	(a)		(b)	
Variable	Value	*t*-Ratio	Value	*t*-Ratio
Constant	−2.63	1.70	−2.64	1.76
Per capita GDP	0.26	2.12	0.27	1.96
Gini index	−0.75	2.40	−0.76	2.47
Taxes	0.73	2.76	0.72	2.83
Education	0.33	2.03	0.32	1.70
Openness rate	0.12	1.04	0.12	1.04
Middle East & North Afr.	−0.59	2.49	−0.59	2.49
Europe & Central Asia	−0.69	4.81	−0.69	4.73
Adjusted R^2	0.80		0.80	
No. of observations	78		78	
Infra-identification test: statistic Kleibergen-Paap (χ^2 value (6))	14.88 (12.59)		14.00 (12.59)	
Statistic J (χ^2 value (5))	10.82 (11.07)		10.89 (11.07)	
Endogenous	Per capita GDP 2004, Gini Index, taxes and openness rate		Per capita GDP 2004, Gini Index, taxes, education and openness rate	
Instruments	Per capita GDP 1990, (Per capita GDP 1990)^2, fuel exports, ethnic fragmentation, population and regional dummies		Per capita GDP 1990, (Per capita GDP 1990)^2, fuel exports, education 1990, ethnic fragmentation, population and regional dummies	

Source: Alonso and Garcimartín (2013, p. 217)
(a) Education is considered exogenous. (b) Education is considered endogenous
Instrumental variables. Robust estimates

Measuring Governance: A Tentative Proposal

The discussion above about the functions that institutions should fulfil drove us to identify four criteria to evaluate institutional quality. The empirical analysis, even though not conclusive, supported our view. However, two remarks need to be made: the state is a specific component of a country's institutional framework with a particular role; and governance embraces more actors than the state, including the relationships among state, the private sector, and civil society in the public realm. Therefore, how can the aforementioned criteria for institutional quality inspire the measurement of good governance?

The current literature on measuring good governance poses plenty of difficult dilemmas. Some analysts think that governance should be measured by what the state produces (outputs and outcomes) (Rotberg 2014; Andrews 2014), while others argue that governance must be defined by looking at how

governments operate (capacity and autonomy of the State: Fukuyama, 2014; or impartiality: Rothstein 2011). There are authors who hold that governance as implementation should be separated from the normative ends that governments are meant to serve (Fukuyama 2013), while others consider that the way in which this normative mandate is fulfilled by governments should necessarily be part of the analysis (Rothstein 2011). Finally, some researchers operate by collecting very disaggregated information focused on just one outcome in one field to properly measure governance (Andrews et al. 2010; Andrews 2014), while others support a more "macro-systemic" approach (Rolland 2004).

It is very difficult to settle on a consensual opinion between these opposing positions. Here, however, we will set out our view in order to explain our methodological choices.

First, it is very difficult to see how the quality of governance can be assessed without *normative implications* related to the functions (or attributes) that good governance has to fulfil, particularly if the fulfilment of these functions conditions the nature of relationships among social actors and affects the capacity of institutions to make and enforce rules. It is nearly a contradiction to define "good governance" without referring to those attributes of governance that we consider to be "good". In any case, it is important to avoid the tautological approach that would result from defining good governance in terms of the goals that we previously characterised as representing success. For this reason, we talk about "general functions or attributes" (such as stability, adaptability, capacity, or credibility) as being different from "transformative goals". In the same vein, we distinguish governance attributes from concrete policy choices.

Second, there are good reasons for measuring good governance in terms of *procedures, attributes, or general functions* of governance rather than in terms of their outputs and outcomes. As historical experience reveals, bad institutional frameworks can temporarily produce good outcomes, and vice versa. Moreover, some important outputs (and outcomes) are not simply the consequence of public actions, but the result of a complex and poorly understood array of factors related to a country's level of development, historical inertia, or international circumstances, among others. For example, with a limited ability to define clear causal relations it would be difficult to associate outputs in social services with a measure of good governance. Even if econometric estimates are controlled by country's income—as in Andrews (2008)—the link between outputs and quality of governance is questionable (and it leads to paradoxical results, with better governance in health in Pakistan than in the United States, for example; see Andrews et al. 2010).

Third, the institutional framework is not a juxtaposition of autonomous institutions, but a *hierarchical and connected structure* of institutions that operate at different levels. It is difficult to change one institution if other related structures are not also altered. There are strong complementarities and substitutability among different institutions. As a consequence, a functionalist approach based on identifying specific outputs in very specific fields to measure good governance can be inappropriate. It seems better to adopt a systemic approach that considers groups of highly connected institutions.

Finally, the aim of this effort to give empirical support to the concept of governance should be to help to clarify the *link between indicators and reform action*. Data should be useful for making a suitable diagnosis of the weaknesses and shortcomings of governance and in assisting policymakers with identifying and evaluating specific policy decisions. In sum, the criteria finally chosen should easily be transformable into sound reliable operational data.

All these assumptions could be acceptably met using the "good governance" criteria proposed by Fukuyama (2013) (capacity and autonomy), as well as that suggested by Rothstein (2011, 2013) (impartiality). However, in the first case an excessively state-centric conception of governance is adopted and consequently good governance is narrowly assessed in relation to the administrative and executive capacity of the state. Important components of our definition of governance—related to the relations between the state and its citizens, such as stability and credibility—remain out of focus. In the second case, even if impartiality is an important element for building credible institutions, we agree with Fukuyama (2013) that by itself it is not a sufficient metric for measuring good governance. In an attempt to overcome these shortcomings and offer a more comprehensive approach, as a metric of good governance we suggest the four attributes identified before as criteria for institutional quality (capacity, predictability, credibility, and adaptability).

Before applying these criteria in operative terms, we need to identify groups of interconnected institutions, taking into account the general roles of the state. Let us give an example regarding the state's economic functions (other functions could be considered in a more complete assessment). Table 4.3 illustrates this tentative approach. The main systemic fields of functions of market-supporting institutions could be the following six (another taxonomy with a similar approach is offered by Rodrik 2007): (i) regulating basic economic rights, (ii) promoting market competition, (iii) coordinating and enlarging markets, (iv) correcting externalities, (v) promoting macroeconomic stability and growth, and (vi) promoting social cohesion and managing distributive conflicts.

Table 4.3 Economic roles of the state

Area	Variable	Capacity	Predictability	Adaptability	Credibility
Basic economic rights	Appropriate institutions for defining and defending property rights				
	Appropriate institutions for ensuring safe economic transactions				
Market competition promotion	Appropriate institutions for regulating market access				
	Appropriate institutions for developing and enforcing competition rules				
	Appropriate institutions for defending consumer rights				
	Appropriate regulation to preserve transparency and competition in public procurement				
	Markets open to international competition				
	Appropriate regulation on corporative governance				
Market coordination and market enlargement	Appropriate investment incentives				
	Appropriate policy for promoting internationalisation				
	Appropriate regulation of labour markets (including labour mobility)				
	Appropriate institutions for promoting innovation and technological capacities				
Correction of externalities	Appropriate public infrastructure				
	Appropriate education system				
	Appropriate health system				
	Appropriate institutions for protecting the environment				
Macroeconomic stability and growth	A transparent and stable regulatory framework				
	An appropriate tax system				
	Appropriate institutions for preserving economic stability				
	Appropriate institutions for regulating financial markets				
Social cohesion and management of distributive tensions	Appropriate institutions for social protection				
	Appropriate institutions for the inclusion of the poor				
	Appropriate institutions for correcting regional disparities				
	Appropriate institutions for channelling social tensions and adequate representation of the different social sectors				
	Appropriate institutions for defending collective interests and promoting social agreements				

Source: elaboration by the authors

Each of these fields has different components in accordance with the roles that the state should play and the institutions that are required to govern them. We can apply the four criteria of institutional quality to each component. Different indicators related to each of our four criteria for institutional quality can be identified to assess governance in each component (see Table 4.3). It is important to use disaggregated information about the different components of the roles of the state, as the quality of governance can be different depending on the component considered. However, identifying adequate indicators for each cell of the table is obviously a complex task.

This proposal can be illustrated with an example. We identify one indicator for each cell in the "Social cohesion and management of distributive conflicts" field. It is important to test the extent to which these indicators reliably reflect the components (rows) and criteria (columns) previously defined. A good way to do this is through comparative analysis, applying the indicators to different countries and checking the ability they have to reach an appropriate diagnosis. We admit that this is not a simple task in practice; our example (Table 4.4) is a mere illustration.

As an empirical exercise, we have applied this approach to two clearly comparable countries: Uruguay and Panamá—two Latin American countries with similar populations (Panamá, 3.8 million; Uruguay, 3.4 million, both in 2013) and GDPs per capita in PPP (Panamá, US$ 19,894; Uruguay, US$ 19,945, in the same year) and with economic structures mainly based on services. We have tried to fill every cell in the above table using indicators from the same sources for both countries, although unfortunately information was not available in all cases (see the Annex for the statistical sources).

In order to ensure comparability and to obtain a clearer picture of governance quality, all the variables have been normalised (as explained below) to range between 0 and 1, where 0 (1) is the lowest (highest) value of each variable in the Latin American region. In other words, the different values in the table above reflect the scores obtained by Panamá and Uruguay within the Latin American context. Table 4.5 shows the result of this exercise.

The total scores for each aspect of governance quality are shown in Table 4.6. Uruguay shows higher values for both the total overall score and each particular aspect of institutional quality. However, while in terms of efficiency the results for Uruguay are about 30% higher than those for Panamá, in the predictability and especially in the credibility fields, these differences are much more pronounced: 54% and 75%, respectively. Therefore, if the quality of social cohesion institutions in Panamá is to be increased, efforts should be particularly focused on improving the predictability and credibility of institutional arrangements.

Table 4.4 Social cohesion and management of distributive tensions: indicators

Variable	Capacity	Predictability	Adaptability	Credibility
Appropriate institutions for social protection	Coverage of social security (% of population)	Stability of social expenditure over the last ten years (% of GDP)	Unemployment rate in the first and second quintile (%)	Social expenditure (% of public budget)
Appropriate institutions for the inclusion of the poor	% of population under the national poverty line	Stability of public expenditure oriented to fighting poverty over the last ten years	% of school enrolment in the first quintile	Fighting poverty as priority in national development strategies
Appropriate institutions for correcting regional disparities	Territorial disparities	Changes in regional distribution of public expenditure	Coefficient of convergence among regions in the last decade	Regional representatives in the national parliament
Appropriate institutions for channelling social tensions and adequate representation of the different social sectors	Number of strikes per year	Changes in labour regulation over the last ten years	Existence of economic and social councils	Membership of trade unions (%)
Appropriate institutions for defending collective interests and promoting social agreements	General strikes and public demonstrations against the government	% of government legislative initiatives passed by the parliament	Existence of official projections for national development	Number of advisory councils open to social participation

Source: elaboration by the authors

Table 4.5 Social cohesion and management of distributive tensions: indicators for Panamá and Uruguay

	Capacity		Predictability		Adaptability		Credibility	
Appropriate institutions for social protection	**Coverage**		**Social expenditure stability**		**Unemployment**		**Social expenditure**	
	Panamá	Uruguay	Panamá	Uruguay	Panamá	Uruguay	Panamá	Uruguay
	0.25	0.69	0.93	1	0.38	1	0.25	0.87
Appropriate institutions for the inclusion of the poor	**Poverty**		**Stability of public expenditure oriented to fighting poverty**		**% of school enrolment**		**Poverty reduction**	
	Panamá	Uruguay	Panamá	Uruguay	Panamá	Uruguay	Panamá	Uruguay
	0.70	0.95	NA	NA	0.60	1	0.67	1
Appropriate institutions for correcting regional disparities	**Territorial disparities**		**Changes in regional distribution of public expenditure**		**Coefficient of convergence among regions in the last decade**		**Regional representatives in the national parliament**	
	Panamá	Uruguay	Panamá	Uruguay	Panamá	Uruguay	Panamá	Uruguay
	0.60	1	NA	NA	NA	NA	0	0
Appropriate institutions for channelling social tensions and adequate representation of the different social sectors	**Number of strikes per year**		**Trust in political institutions**		**Sub-employment**		**Affiliates to labour unions**	
	Panamá	Uruguay	Panamá	Uruguay	Panamá	Uruguay	Panamá	Uruguay
	0.99	0.82	0.35	1	0.99	0.87	0.44	0.68
Appropriate institutions for defending collective interests and promoting social agreements	**Institutionalised democracy and participation**		**Regime stability**		**Existence of official projections for national development**		**Institutionalised constraints on decision-making powers**	
	Panamá	Uruguay	Panamá	Uruguay	Panamá	Uruguay	Panamá	Uruguay
	0.83	1	0.14	0.19	1	1	0.67	1

Source: elaboration by the authors

Table 4.6 Total scores for Panamá and Uruguay

	Capacity	Predictability	Adaptability	Credibility	Total
Panamá	3.37	1.42	2.97	2.03	9.79
Uruguay	4.46	2.19	3.87	3.55	14.07
Ratio (Pn/Ur)	1.32	1.54	1.30	1.75	1.44

Source: elaboration by the authors

Our four criteria for institutional quality however have to be considered complementary factors and therefore potential trade-offs among them can emerge. For example, improvements in "institutional predictability" hence could imply costs in terms of "institutional adaptability". As a consequence, aggregate indicators should be built in columns rather than in rows for each area considered, that means separate results for each institutional quality criterion should be considered. In addition, the problem of assigning weights to the individual indicators in each column needs to be further explored.

Policy Implications

As stated above, empirical research should not be limited to investigating whether or not institutions and governance play a relevant role in development, but it must also aim to identify the determinants of institutional quality so that policies addressed at building better governance can be developed. To do this, we must first have a clear picture of what "good institutions" are, and in our view the picture underlying most institutional quality indicators is too narrow. They usually assume that institutions are mere substitutes for markets and as a result the quality of the former can be measured by the same criteria as those used to measure the quality of the latter: in essence, limited intrusion into agents' behaviour and flexibility. Our approach, on the contrary, is based on the idea that institutions are not only created to introduce flexibility or market incentives for social behaviour but also to provide some regularity in human conduct, to reduce uncertainty and risk in social interaction, and to shape collective action.

This conception of institutions not only has different implications regarding institutional quality indicators, but it also differs from the usual approach with respect to policy recommendations. Generally, in the literature, "ideal institutions" would seem to be defined as those of successful countries, and therefore recommendations of transplants emerge. However, according to our proposal things are more complex. The analysis of the four dimensions of institutions developed above implies that policies aimed at building better

institutions must pay attention not only to their efficiency (static and dynamic) but also to their predictability (stability) and credibility (legitimacy). This conception has three major policy implications.

First, as highlighted in the example provided in Table 4.4, the four dimensions of institutional quality must be analysed separately, and each of them includes a number of areas and factors. Policy responses should be based on such an assessment. In some cases policy action should be oriented to giving the existing institutions more stability (and certainty), which entails operating with reforms at the margin (rather than with radical changes). In other cases, the problem may be a need for more credible institutions, which requires voice, transparency, enforcement, or a better redistribution of collective action benefits. Sometimes the limitation is the level of static efficiency of certain institutions, which indicates a need to review procedures, reduce discretionary decision making, and fight the activity of rent-seekers. Finally, problems can emerge of institutions weakly adapted to a new social reality, requiring changes in the design of incentives. Each dimension defines a different line of policy priorities.

Second, frictions, conflicts, and trade-offs can emerge among these four dimensions; for example, predictability can be increased at the expense of reducing dynamic efficiency; and credibility can affect static efficiency. It is important to identify the problem that should be tackled, but also the effects that the measures to be adopted can have along the other dimensions.

Finally, institutional quality policies are in most cases country- and time-specific, depending on the weights of the different dimensions in each case, the areas to be prioritised, and the potential trade-offs among them. For example, dynamic efficiency can be more relevant in the early stages of development and predictability after episodes of high sociopolitical instability. There are therefore no universal policy recommendations (Andrews 2013).

This last comment can be extended in order to illustrate the policy implications of our approach. We have defined four criteria of institutional quality, but international experience shows that the relevance of each dimension depends on the country's stage of development. This is very important when countries with episodes of intense economic growth are analysed: Taiwan, Korea, or Singapore, for example, but also Spain or Brazil in the 1960s. The role played by the state during the economic transformation of these countries was crucial. An ample array of studies have analysed the features that characterise this kind of state, named "Developmental States" (among others by Johnson 1982; Evans 1995; Woo-Cummings 1999). This is not the place to discuss the insights of this literature, but in the terminology of our approach the two dimensions that were prioritised by these states were stability and

dynamic efficiency. They were authoritarian (legitimacy dimension) and corrupt (at the cost of the static efficiency dimension) states, but they provided security to their populations (stability) and intense growth through continuous changes in their productive specialisation (dynamic efficiency). After reaching a certain level of development, the relevance of the other two dimensions began to increase (and their populations demanded more democracy, more transparency, and less corruption).

In a more general framework, we could draw some conclusions about the path of institutional change that usually accompanies the development process. Developing countries tend to have institutional frameworks with deficiencies in all four of our dimensions. Donors try to shift these frameworks towards the parameters of developed countries (with better scores in all dimensions) through institution building programmes. However, the path should not necessarily be that of an isomorphic change (in all the four dimensions). Quite to the contrary, institutional reform programmes may require concentration on some specific dimensions as prior conditions for further reforms in the others (as the experience of the Developmental States suggests). The process is far from linear and comes close to a mechanism of interlinked imbalances (as Hirschman 1958, suspected). This is an important conclusion for the orientation of appropriate institutional quality reform programmes.

Conclusions

Five main conclusions emerge from our analysis. First, more reliable and appropriate indicators are needed to understand and promote good governance. Over the past two decades, significant effort has been dedicated towards building new databases and datasets of indicators to measure institutional quality and good governance, yet the situation is still far from satisfactory. Available indicators suffer from various shortcomings and offer limited support for steering diagnostic activities and policy action.

Second, before developing new indicators we should define more carefully the theoretical basis used to build them and to select variables. Most databases and datasets have been built on the collection of valuable information about different aspects of good governance. The reflection on assumptions about the relationship between those indicators and the theoretical variables that condition good governance has yet been limited. This gap needs to be closed. Hence, more conceptual debate is needed to build better indicators of good governance.

Third, since governance is defined within the framework of the set of relations among social actors established to articulate collective responses to common problems, the institutional framework is crucial to the understanding of governance. Therefore, governance indicators can be based on criteria related to institutional quality, adapting them to the different functions of the state.

Fourth, and in accordance with this approach, we suggest four criteria of institutional quality to be applied to measuring good governance: capacity, predictability, adaptability, and credibility. These criteria have been confirmed as empirical determinants of institutional quality and might guide the search for indicators of good governance in each of the fields in which the state operates. They yet have to be considered complementary factors and, therefore, should be measured separately given that a state function can score high in one dimension, but low in another. Such disaggregate information by fields and function of the state as well as criteria of good governance could therefore be a useful support for tailoring policy reform.

Finally, a note of caution is needed. Our proposal is work in progress that requires further discussion and testing. The indicators and data discussed above have been offered only as an illustration of the way in which our approach could operate. More work is needed for a better grasp of the whole picture of what good governance means.

Annex. Variables: Definition and Sources

All variables have been normalised to range between 0 and 1 by computing $Xi = \frac{xi - xmin}{xmax - xmin}$ or $= 1 - \frac{xi - xmin}{xmax - xmin}$, where i stands for Panamá or Uruguay, and $x_{\max}$ and xmin refer to the highest and lowest values of the relevant variable for Latin American countries.

Coverage:	Percentage of population participating in social protection and labour programmes (all social insurance) (last available). Source: ASPIRE, The World Bank
Social expenditure stability:	Social expenditure as a percentage of GDP coefficient of variation, 2000–2009. Source: ECLAC

Unemployment:	Long-term unemployment: unemployed 12 months or more as a percentage of total unemployment; 2009–2012. Source: ILO
Social expenditure:	Social expenditure as a percentage of total public expenditure (last available). Source: own calculations based on ECLAC
Poverty:	Poverty headcount ratio at national poverty levels (% of population) (last available); Source: The World Bank
Percentage of school enrolment:	Primary school completion rate (bottom 20%/ top 20%); 2005. Source: ECLAC
Poverty reduction:	Annual reduction of poverty headcount ratio (2000–2013). Source: own calculations based on The World Bank
Territorial disparities:	Coefficient of variation of regional GDP per head. Source: based on Garcimartín, C., Rivas, L., Díaz de Sarralde, S. and Alonso, J. A (2009): ¿Es viable un Fondo de Cohesión Social Iberoamericano? Georgetown University- Universia VOL. 3 NUM. 1
Regional representatives in the National Parliament:	Neither Uruguay nor Panamá has regional representatives in their parliaments
Number of strikes per year:	Strikes and lockouts, 2008. Source: ILO
Trust in political institutions:	Percentage of population who trust political institutions and the state; 2013. Source: ECLAC
Sub-employment:	Sub-employment rate (last available). Source: ECLAC
Membership of labour unions:	As a percentage of employees, 2010. Source: ILO
Institutionalised democracy and participation:	Polity variable in the Polity IV Project. 2013
Regime stability:	Durable variable in the Polity IV Project. 2013
Institutionalised constraints on decision-making powers:	Xconst variable in the Polity IV Project. 2013

References

Acemoglu, D., Johnson, S., & Robinson, J. A. (2001). The Colonial Origins of Comparative Development: An Empirical Investigation. *American Economic Review, 91*(5), 1369–1401.

Acemoglu, D., Johnson, S., & Robinson, J. A. (2002). Reversal of Fortunes: Geography and Institutions in the Making of the Modern World Income Distribution. *Quarterly Journal of Economics, 117*(4), 1231–1294.

Acemoglu, D., Johnson, S., & Robinson, J. A. (2005). Institutions as the Fundamental Cause of Long-Run Growth. In P. Aghion & S. Durlauf (Eds.), *Handbook of Economic Growth* (pp. 358–472). Amsterdam: Elsevier/North-Holland.

Alesina, A., Devleeschauwer, A., Easterly, W., & Kurlat, S. (2003). Fractionalization. *Journal of Economic Growth, 8*(2), 155–194.

Alonso, J. A. (2011). Colonisation, Institutions and Development: New Evidence. *Journal of Development Studies, 47*(7), 937–958.

Alonso, J. A., & Garcimartín, C. (2008). *Acción Colectiva y Desarrollo. El Papel de las Instituciones*. Madrid: Editorial Complutense.

Alonso, J. A., & Garcimartín, C. (2013). The Determinants of Institutional Quality. More on the Debate. *Journal of International Development, 25*(2), 206–226.

Andrews, M. (2008). The Good Governance Agenda: Beyond Indicators without Theory. *Oxford Development Studies, 36*(4), 379–407.

Andrews, M. (2013). *The Limits of Institutional Reform in Development. Changing Rules for Realistic Solutions*. Cambridge: Cambridge University Press.

Andrews, M. (2014). *An Ends-Means Approach to Looking at Governance (Center for International Development Working Paper 281)*. Cambridge, MA: Harvard University.

Andrews M., Hay, R., & Myers, G. (2010). *Governance Indicators Can Make Sense. Under-Five Mortality Rates Are an Example* (HKS Faculty Research Working Paper Series, RWP10-015). Cambridge, MA: John F. Kennedy School of Government, Harvard University.

Aoki, M. (2001). *Towards a Comparative Institutional Analysis*. Cambridge: MIT Press.

Arndt, C., & Oman, C. (2006). *Uses and Abuses of Governance Indicators*. Paris: OECD.

Booth, D., & Cammak, D. (2013). *Governance for development in Africa. Solving collective actions problems*. London: ZED Books.

Buchanan, J. M., & Tullock, G. (1962). *The Calculus of Consent*. Ann Arbor: University of Michigan Press.

Chong, A., & Zanforlin, L. (2000). Law Tradition and Institutional Quality: Some Empirical Evidence. *Journal of International Development, 12*(8), 1057–1068.

Easterly, W., & Levine, R. (2003). Tropics, Germs, and Crops: How Endowments Influence Economic Development. *Journal of Monetary Economics, 50*(1), 3–39.

Easterly W., Ritzan J., Woolcock, M. (2006). *Social Cohesion, Institutions, and Growth* (Center for Global Development Working Paper 94, August). Washington: Center for Global Development.

Evans, P. (1995). *Embedded Autonomy: States and Industrial transformation*. Princeton: Princeton University Press.

Evans, P. (2004). Development as Institutional Change: the Pitfalls of Monocropping and the Potential of Deliberation. *Studies in Comparative International Development, 38*(4), 30–52.

Fukuyama, F. (2011a). *The Origins of the Political Order: From Prehuman Times to the French Revolution*. New York: Farrar, Straus and Giroux.

Fukuyama, F. (2011b). *Political Order and Political Decay: From the Industrial Revolution to the Globalization of Democracy*. New York: Farrar, Straus and Giroux.

Fukuyama, F. (2013). *What is governance?* (Centre for Global Development Working Paper, 314). Washington, DC: Center for Global Development.

Fukuyama, F. (2014). *Political Order and Political Decay. From Industrial Revolution to Globalization and Democracy*. New York: Farrar, Straus and Giroux.

Furubotn, E. G., & Richter, R. (2005). *Institutions and Economic Theory. The Contribution of the New Institutional Economics*. Ann Arbor: The University of Michigan Press.

Glaeser, E. L., La Porta, R., López-de-Silane, F., & Shleifer, A. (2004). Do Institutions Cause Growth? *Journal of Economic Growth, 9*(3), 271–303.

Greif, A. (2006). *Institutions and the Path to Modern Economy. Lessons from Medieval Trade*. Cambridge: Cambridge University Press.

Greif, A., & Kingston, C. (2011). Institutions: Rules or Equilibria? In N. Schofield & G. Caballero (Eds.), *Political Economy of Institutions, democracy and voting* (pp. 13–43). Heidelberg: Springer-Verlag.

Hall, R. E., & Jones, C. I. (1999). Why do Some Countries Produce So Much More Output per Worker Than Others? *Quarterly Journal of Economics, 114*, 83–116.

Hirschman, A. O. (1958). *The Strategy of Economic Development*. New Haven: Yale University Press.

Islam, R., & Montenegro, C. (2002). *What Determines the Quality of Institutions? World Development Report: Building Institutions for Markets*. Washington, DC: World Bank.

Johnson, C. (1982). *MITI and the Japanese miracle. The Growth of Industrial Policy, 1925-1975*. Stanford: Stanford University Press.

Kaufmann, D., Kraay, A., & Mastruzzi, M. (2007). *The Worldwide Governance Indicators Project: Answering critics* (World Bank Policy Research Paper 4149). Washington: World Bank.

Kooiman, J. (2003). *Governing as Governance*. London: Sage Publications.

La Porta, R., López de Silanes, F., Shleifer, A., & Vishny, R. W. (1999). The Quality of Government. *Journal of Law, Economics and Organization, 15*(March), 222–279.

Levi-Faur, D. (2013). What is Governance? [Blogpost]. The Governance Blog. Retrieved from https://governancejournal.net/2013/03/10/levi-faur-on-what-is-governance

Moe, T. M. (1990). Political Institutions: The Neglected Side of the Story. *Journal of Law, Economics, and Organization, Special Issue, 6*, 23–53.

Moore, M. (2004). Revenues, State Formation and the Quality of Governance in Developing Countries. *International Political Science Review, 25*, 297–319.

North, D. (1990). *Institutions, institutional change and economic performance*. Cambridge: Cambridge University Press.

North, D. (2005). *Understanding the process of economic change*. Princeton: Princeton University Press.

Rigobon, R., & Rodrik D. (2004). *Rule of Law, Democracy, Openness, and Income: Estimating the Interrelationships* (National Bureau of Economic Research Working Paper 10750). Cambridge, MA: National Bureau of Economic Research.

Rodrik, D. (2007). *One Economics, Many Recipes. Globalization, Institutions and Economic Growth*. Princeton: Princeton University Press.

Rodrik, D., Subramanian A, & Trebbi F. (2002). *Institutions Rule: The Primacy of Institutions over Geography and Integration in Economic Development* (International Monetary Fund Working Paper 02/189). Washington: International Monetary Fund.

Rolland, G. (2004). Understanding Institutional Change: Fast-Moving and Slow-Moving Institutions. *Studies in Comparative International Development, 38*(4), 109–131.

Rotberg, R. I. (2014). Good Governance Means Performance and Results. *Governance, 27*(3), 511–518.

Rothstein, B. (2011). *The Quality of Government: Corruption, Social Trust, and Inequality in International Perspective*. Chicago: Chicago University Press.

Rothstein, B. (2013). *The Three Worlds of Governance: Arguments for a Parsimonious Theory of Quality of Government* (The Quality of Government Institute Working Papers Series 2013:12). Göteborg, Sweden: The Quality of Government Institute, University of Göteborg. Retrieved from http://qog.pol.gu.se/digitalAssets/1456/1456679_2013_12__rothstein.pdf

Scott, W. R. (2008). *Institutions and Organizations. Ideas and Interest*. London: Sage Publications.

Solow, R. M. (1956). A Contribution to the Theory of Economic Growth. *Quarterly Journal of Economics, 70*(1), 65–94.

Suchman, M. C. (1995). Managing Legitimacy: Strategic and Institutional Approaches. *Academy of Management Review, 20*, 571–610.

Thomas, M. A. (2009). What Do the Worldwide Governance Indicators Measure? *European Journal of Development Research, 22*(1), 31–54.

Woo-Cummings, M. (1999). *The Developmental State*. Ithaca, NY: Cornell University Press.

5

The Creative Disorder of Measuring Governance and Stateness

Debora Valentina Malito

Introduction[1]

Measures of governance have grown substantially in number and typologies over recent decades. The concept of governance has attracted considerable attention when the end of the Cold War enabled the diffusion of the neoliberal political order. New political structures and actors acquired authority and exercised multi-forms of power. As pointed out by Barnett and Duvall (2005), indirect forms of power gained relevance in the context of the liberal global economy. Productive power concerns the production of subjectivity, while institutional forms of power is about the design of formal and informal rules and practices of governance through which institutions orient power relationships (Barnett and Duvall 2005, p. 3). The term governance gained hence prominence as a theoretical approach and concept grasping different aspects. Yet, what does governance exactly mean? Three main lines of argumentation should be considered.

First, to differentiate governance from government and/or governing, many scholars emphasised the type of actors on whom steering functions are vested. Rosenau pointed out that governance is by definition "always effective in performing the functions necessary to systemic persistence" (Rosenau and Czempiel 1992). Governments succeeded indeed in their governance functions only when the formal state authority is capable of

D.V. Malito (✉)
University of Cape Town, Cape Town, South Africa

D.V. Malito et al. (eds.), *The Palgrave Handbook of Indicators in Global Governance*,
https://doi.org/10.1007/978-3-319-62707-6_5

enforcing decision-making. Governance is indeed a form of collective action, "inherently comprehensive of both government and civil society" (Lynn 2013, p. 13). Focusing on networks, many scholars explored the role of corporations (Wilkes 2004) or private actors. For Craig, global governance concerns a scheme made of global regulations sustained by a transnational class of social forces (Murphy 2000), where private sector markets and non-governmental organisations exercise a new form of legitimate power. Others emphasised the plurality and multi-level of actors involved, in which classical authorities do not completely disappear. It "is not what the state does that is different, but *how* it does it" (Richards and Smith 2002).

Secondly, focusing on aspects of authority, scholars claimed that the peculiarity of governance lies on the formation of different forms of authority. The idea of a global authority is built upon a pattern of legal pluralism that attributes to the contemporary mode of governance a hegemonic, consensual character. While Thomas Weiss uses the concept of plurality (Weiss and Wilkinson 2013, p. 167), Cluter claims that a form of "common good" operates dialectically as the "common sense" of contemporary global capitalism (Cutler 2013, p. 719). For Soderbaum, the creation of common values is not just a "procedural and technical exercise", but rather a "political attempt to construct an imagined community between states" (Soederberg 2004, p. 142). Indispensable to the transmission of such a global form of authority is what Lake defines a social relationship and mutual agreement established between the ruler and the ruled (Lake 2009, p. 332). By elaborating on the Gramscian conception of hegemony, Robert Cox claims that a transnational form of consensus regulates governance practices, where "the more powerful must act in such a way that their interests have the appearance of a general interest" (Cox 1980, p. 376). The governance function is performed by a "nébuleuse, a loose elite network of influential and agencies, sharing a common set of ideas" (Cox 1997, p. 61). This nébuleuse of governing actors generates "a common economic ideology and inject this consensual outcome into national processes of decision-making" (Cox 1997, pp. 59–61).

Thirdly, the concept of governance became particularly prominent within EU studies. As an antipode to classical notions of government, governance became a heuristic and technical device defining the changing nature of the state under conditions of supranational political integration, that is, the emergence of multi-level, co-operative decision-making structures. Table 5.1 summarises a way of differentiating between government and governance, considering differences on politics, policy, and polity. This technocratic

Table 5.1 Government vs. governance

	Government State vs. market resp. society	Governance State, market, and networks as complementary forms of steering
Polity	Focus on the state Majoritarian democracy and hierarchy as most important institutions	Institutional structure, which combines elements of hierarchy, negotiation systems, and competition mechanisms Networks
Politics	Competition between political parties for acquisition of power and between interest groups to gain influence Conflict regulation by decision of responsible state organs and enforcement of official decisions	Conflicts between governing/leading and governed/affected actors Steering and co-ordination within institutional steering systems Negotiations between state and/or societal actors Adaptation of institutional steering systems
Policy	Legislation (order and prohibition) Distribution of public goods	Agreement (within networks and communities), compromises, barter Co-production of collective goods Network management Institution building (management of institutional change)

Source: Umbach 2007, p. 37 (translated version of Benz 2004, p. 21)

"governance turn" (Rosamond 2000, p. 110) was also accompanied by a strong focus on aspects such as administrative capacity, bureaucratic functioning, and institutional capacity.

Because of the ubiquity and slippery nature (Cox 1997, pp. 59–61) of the governance paradigm, the term gained various meanings and functions. Like the concept of sustainable development, governance seemed to turn into a catch-all notion for political processes and a "fetish" (Levi-Faur 2012) used to refer to everything and anything (Finkelstein 1995). Over the life cycle of the concept, a new theoretical debate re-shaped the course of the governance debate. After a decade in which the governance paradigm challenged classical conceptions of state authority, the sovereign state re-emerged as key conceptual reference point (Bhuta 2012). With the escalation of the Global War on Terror, the international community began to address the problem of "state weakness". The "collapse" or "failure" of states was considered a crisis of governance producing threats to the global security. Such insecurity was semantically used to reinforce the search for good governance instruments. As a result, many practitioners and scholars directly related state weakness with the discourse on governance. In view of these developments and challenges, the governance-stateness nexus became more complicated.

Yet, the creative disorder in the way of defining and measuring governance is not incidental. It rather reflects the hybridity and pervasiveness of the neoliberal paradigm of governance. The epistemic confusion, overlap, in the terminology employed to some extent derives from the climate of transition in which the free market model of governance emerged. With the crisis of the Bretton Woods system, a new international political order emerged. While the distribution of power shifted with the end of the Cold War, and nation-states remained the gatekeeper of de jure sovereignty, is the way of exercising power that changed with the neoliberal rise of the 1970s.

The governance discourse hence became a functional reaction to a crisis of governability (Levi-Faur 2012); stimulated by the need of creating a new, pervasive, and pragmatic approach to capitalism, marked by the "destatisation" of the political system (Jessop 1997). The term governance emerged as the catch-all notion for all the economic and political processes happening under liberalisation and free market policy prescriptions. Here, new forms of power and authority are exercised via decentralised or cooperative mechanisms, based on soft forms of persuasion and attraction policies (Nye 1990). Yet, governance represents also an important asset of the modernisation discourse. According to Fine (2009, p. 12):

> in place of the amorphous but at least single notion of modernisation, there has been the proliferation of developmental terminology, from good governance through to corruption, each element of which has to be critically unpicked across rhetoric, etc., to make any sense of what is being said or, possibly more important, being done.

By following these insights, this chapter maps the literature on governance and stateness indicators. It aims to explore strengths and weaknesses of the current debate. It claims that the proliferation of indicators can only be understood against the conceptual hybridity and indeterminacy in which the notions of governance and stateness have in recent decades become increasingly entangled. To frame and explore such "creative disorder" (Kooiman 2003), the chapter provides an overview on the production of governance and stateness indicators. It then reconstructs first the conceptual meaning of governance and stateness. It explores potential normative demands and policy prescriptions linked to their production. The chapter then analyses the conceptual construction of both governance and stateness indicators, considering three different measures (The Rule of Law Index, the Sustainable Governance Indicators and the State Fragility Index). The chapter concludes by summing up the relevance of exploring the conceptual and normative context of these measures.

Measurements of Governance

There are few terms in social science and in practice (Bovaird and Löffler 2003, p. 316) that are as vague as the governance concept. Yet, definitions, theories, and measures of governance have still grown substantially in number over the recent decade. The term governance has attracted a pluralist connotation both in the academic and public debate. Scholars, institutions, and stakeholders have provided as many definitions as to turn the term into a mixed bag of notions, aspects, and attributes. Measures of governance have also proliferated during the last decades. Yet, there seems to be little agreement on what exactly governance means.

In order to reconstruct the systematised meaning associated with the concept of governance, Table 5.2 illustrates a set of definitions of governance provided by international organisations. It shows the multiple interpretations of the term governance, in which the core essence of the concept substantially varies from one provider to the other. As a result of this multiplicity, measures have been operationalised as diversely as processes (governance as a tool) and

Table 5.2 Definitions of governance

Institutions	Definitions	Characteristics
EC	"Governance represents both the rules, process, and behaviour that affect the way in which powers are exercised at the European level, particularly as regard to openness, participation, accountability, effectiveness and coherence"(European Commission 2001).	Input (Process and Structure)
WB	Governance concerns "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 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" (Kaufmann et al. 2011, p. 222).	Input (Process and Structure)
OECD	"Governance is the exercise of political, economic and administrative authority necessary to manage a nation's affairs" (OECD 2014).	Input (Process)
UN	"In the community of nations, governance is considered 'good' and 'democratic' to the degree in which a country's institutions and processes are transparent" (United Nations 2014).	Input (as Quality of institutions and Process)

Source: Own compilation

structures (governance as a form), inputs (governance as both structure and process, bureaucratic and administrative capacity), and outputs (governance as policies, outcomes, consequences, and results). In sum, the nébuleuse of governance metrics reflects the nébuleuse of governing actors informing this process.

The European Commission adopts an input side and subjective (Curtin and Wessel 2005) definition that attributes to governance "both the rules, process, and behaviour that affects the way in which powers are exercised at the European level, particularly as regard to openness, participation, accountability, effectiveness and coherence" (European Commission 2001). For the OECD, indeed, governance is more a process, that is, "the exercise of political, economic and administrative authority necessary to manage a nation's affairs" (OECD 2014). The World Bank (WB) adopts a systemic-input dimension, defining governance as the procedural and structural set of "traditions and institutions by which authority…is exercised" (Kaufmann et al. 2011, p. 222). The UN sees governance as a process and emphasises its instrumental function in promoting democracy: "In the community of nations, governance is considered 'good' and 'democratic' to the degree in which a country's institutions and processes are transparent" (United Nations 2014).

Theoretically, governance is understood as a form of rule without formal government (Rosenau and Czempiel 1992), that is, as a set of regulatory mechanisms implemented without the classical tools of regulation. Despite such institutional perspectives, also the academic debate does not agree on whether governance is a process or a structure, neither whether it should be operationalised through input nor output indicators. As summarised by Umbach (2007) in Table 5.3, governance can be defined using either a process-related perspective or one that focuses on structure.

In the process-oriented perspective, governance is understood "as the continuous political process of setting explicit goals for society and intervening in it in order to achieve these goals" (Jachtenfuchs and Kohler-Koch 2004, p. 99). Here, governance is understood as "non-hierarchical co-ordination between public and private actors, regulated self-steering and societal self-steering" (see Börzel 2005, p. 622; Umbach 2009, p. 40). In a second understanding, "governance encompasses the structural dimension of policy-making as a (new) 'form of social order'" (see Börzel 2005, p. 617; Umbach 2009, p. 41). This structure-oriented perspective indeed encompasses a variety of co-ordination and interaction between interdependent political and societal actors (see Umbach 2009, p. 41).

Both the above perspectives have influenced the formation of governance indicators. Yet, the divergence between input and output perspectives repre-

Table 5.3 Governance structure vs. process

	Process-oriented	Structure-oriented
In a broad sense	Hierarchical co-ordination • Authoritative instruction • Majoritarian decision-making	Hierarchy/State/bureaucracy Independent regulatory authorities Supranational institutions
	Non-hierarchical co-ordination between public and private actors • Negotiation • Conviction	Networks Tripartite negotiation systems Public-private partnerships
	Regulated self-steering (in the shadow of hierarchy)	Neo-corporatist negotiation systems Federations, Interest groups
	Societal self-steering	Community/Clan Market (spontaneous order, anarchy)
In a narrow sense "New" modes of governance in networks	Non-hierarchical co-ordination between public and private actors within processes of arguing and bargaining	Networks Tripartite negotiation systems Public-private partnerships

Source: Umbach 2007, p. 37 (translated and amended version of Börzel (2005, p. 622))

sents still a point of vivid discussion within and beyond the academic debate. While input indicators refer to how governance is implemented (including hence both structures and processes), output indicators take into consideration the consequence of governance in a broad array of sectors. In terms of policy steering, over the past two decades, the measurement of governance has become highly significant in judgements about aid allocation to developing countries. While some actors lamented that output indicators provide very little guidance on why a given country is performing well or not, others criticised the attempt to approach governance through the bureaucratic setting of resources invested (Rotberg 2014). However, both strategies serve different purposes (measuring the quality of governance or its systemic functioning). Instead of debating about the legitimacy of one solution over the other one, more attention should be paid on whether their use is appropriate and in uniformity with their focus and scope. For instance, practitioners tend to evaluate the quality of governance by using indicators of the functioning or bureaucratic organisation of the state. Others, who aim at defining governance from the structural or process-oriented perspective, actually use performance indicators to operationalise their understanding of governance.[2]

During the 1990s, the measurement of governance has acquired a rationalising function, aiming to liberate the concept from both theoretical and concep-

tual uncertainties. As a result of the exponential increase of proxies (corruption, rule of law), attributes (good governance), and correlated variables (democracy, development), the term governance has accumulated a deep connotative capacity but only a spurious denotative understanding (Levi-Faur 2012). To summarise this plurality, Table 5.4 presents some of the most important measures of governance provided by international organisations and stakeholders. Two different strategies of measurement will be considered more closely in order to exemplify different approaches to governance indicator production.

Mono-dimensional Measures

Many indicator providers approximate the value of governance through indirect, mono-dimensional proxy measures. While multi-dimensional indices integrate different representations of the same phenomena in the same measure, mono-dimensional measures account for a single aspect or dimension of governance. They include at least three typologies of measures focusing on: the administrative capacity (i.e. the World Justice Project's (WJP) *Rule of Law*), the quality of political institutions (i.e. the Transparency International's *Corruption Perception Index*; CPI), and the relationship between state and society (i.e. Freedom House's *Freedom in the World*).

Freedom in the World, for instance, is an annual index of political rights and civil liberties, launched in 1972 by Freedom House. It represents the first measure of freedom used to capture the level of governance. Since 1995, Transparency International started producing the Corruption Perception Index, one of the most popular and widespread indicators of corruption used as proxy for measuring governance. Moreover, during the 1990s, the discourse on governance frequently merged with the discourse on poverty reduction, development, and anti-corruption. In 1997, the UN Development Programme launched the *Human Development Index,* broadly considered as either a measure of development and human wellbeing.

Many mono-dimensional measures also capture the quality of political institutions. The strengthening of (good) governance has been a key pillar of the democratisation process undergone by post-colonial states during the 1980s (Weiss 2000). Within the UN agenda for democratisation the "right to democratic governance" (Boutros-Ghali 1996) established a sort of a "global guarantee clause" (see Halperin and Lomasney 1993), according to which the international community should sponsor and support the diffusion and consolidation of liberal democracies. In light of these normative targets, the production of governance indicators was indispensable to measure the transition towards modern economic and political institutions (Cooley 2014).

Table 5.4 Measures of governance

Name	Producer	Mono/ multi-dimensional	Focus	AIM	Coverage	No. of variables
Bertelsmann Transformation Index (BTI)	Bertelsmann Foundation	Mono-	Quality of institutions	"The BTI is directed at the normative goal of a market-based democracy...The BTI emphasizes the same values that underlie the European Union's integration process: a representative democracy under the rule of law combined with a socially responsible and sustainable market economy" (Stiftung 2005, p. 4).	129	2
Corruption Perception Index	Transparency International	Mono-	Administrative capacity	"Raising public awareness of corruption" (Lambsdorff 2007).	176	13
Freedom in the World	Freedom House	Mono-	Quality of institutions	"Analyse the challenges to freedom; advocate for greater political and civil liberties; and support frontline activists to defend human rights and promote democratic change" (Freedom House 2014).	2	195
Global Integrity Index	Worldwide Governance Indicators	Multi-	Quality of institutions	"To assess the existence and effectiveness of mechanisms that prevent abuses of power and promote public integrity, as well as the access that citizens have to their government" (Global Integrity 2014).	43	300
Legatum Prosperity Index	Legatum Institute	Mono-	Global wealth and wellbeing	"Assessment of global wealth and wellbeing" that attempt "to understand how we move 'beyond' GDP" ("The 2013 Legatum Prosperity Index" 2014).	142	89

(*continued*)

Table 5.4 (continued)

Name	Producer	Mono/ multi-dimensional	Focus	AIM	Coverage	No. of variables
Sustainable Governance Indicators	Bertelsmann Foundation	Multi-	Sustainability	"SGI thus targets the spectrum of those individuals who formulate, shape and implement policies, from political decision-makers in centres of government and the democratic institutions of the OECD and EU states, to representatives of civil society and international organizations, to scholars and interested citizen" (Schraad-Tischler, and Seelkopf 2014, p. 2).	41	67
The Country Policy and Institutional Assessment	World Bank	Multi-	State fragility	"The quality of a country's present policy and institutional framework. 'Quality' refers to how conducive that framework is to fostering poverty reduction, sustainable growth and the effective use of development assistance" (The World Bank Group 2011, p. 1).	78	16
The Rule of Law	World Justice Project	Mono-	Administrative capacity (effective exercise of authority)	"The WJP Rule of Law Index is intended for a broad audience of policy-makers, civil society, practitioners and academics, and aims at identifying strengths and weaknesses in each country under review and at encouraging policy choices that advance the rule of law" (World Justice Project 2014, p. 188).	99	43
Worldwide Governance Indicators	World Bank	Multi-	Quality of national governance	"Assessing a measure of governance originally devoted to create cross-country indicators of governance and to establish more effective instruments of government assistance" (The World Bank 2007b).	215	340

Source: Own compilation

However, governance was also instrumental in outlining the path to modernisation and guiding the development aid strategies of donor countries. As result, governance indicators have focused on measuring an array of different, important, aspects of democratic regimes. Since 2003, also the Bertelsmann Foundation provides a *Transformation Index* (BTI) accounting for the political transformation towards democracy (129 countries). The BTI is made up of two indices: the Status and the Management Index. While the first accounts for economic and political transformation, the second is based on indicators of government performance.

Multi-dimensional Measures

The production of governance indicators also resulted in multi-dimensional indices that summarise a large amount of data and information. The World Bank (WB) created one of the first of these multi-dimensional measures, the Country Policy and Institutional Assessment (CPIA). Launched at the end of the 1970s, the CPIA measures the quality of policies and institutional frameworks[3] to guide the investments of the International Development Association, the branch of the World Bank offering loans to developing countries. In 1997, also Freedom House began to calculate a multi-dimensional measure of governance, the Nations in Transit (NIT) index, a measure that aggregates data available on the transition towards democratic institutions undertaken by former communist states in Europe and Eurasia. One of the most widely used multi-dimensional composite measures of governance is the Worldwide Governance Indicators (WGIs) that aggregates different data sources within six dimensions.[4] The World Bank launched the WGI in 1996, when the normative discourse that interlinked development, governance, and democratisation started consolidating. As a result of this new emphasis on a combined perspective, the WGIs aim to provide a more complete and actionable vision of governance. They are, in fact, based on strong policy-making objective, creating the diagnostic tools to steer foreign assistance. The 2013 edition of the WGI is based on 340 variables, and it is available for 215 countries. With this measurement tool, the WB was the first to inaugurate multi-dimensional, composite, measures of governance.[5]

Other institutions have provided multi-dimensional measures using a set of qualities, categories, or attributes (prosperity, sustainable and competitiveness) to better define the nature of governance. In 1980, a UK-based commercial provider launched the International Country Risk Guide

(ICRG), an index of risk assessment composed by 30 indicators and available for 140 countries. Since 2000, also the World Economic Forum (WEF) published the Global Competitiveness Report that, based on two indices (Business Competitiveness Index and the Growth Competitiveness Index), explores the relationship between governance and income. Since 2010, the Legatum Institute provides the Legatum Prosperity Index, a composite index of wellbeing and prosperity, calculated for 142 countries and based on eight clusters of indicators. More recently, the Bertelsmann Foundation launched the Sustainability Governance Indicators that are built on three indices (the Policy Performance Index, the Democracy Index, and the Governance Index) aggregating 67 indicators for 41 countries (edition 2014).

The Semantic Field of Governance

Moving into the semantic field of governance, the relationship between indicators of governance and a set of associated phenomena (such as development, government, state, and corruption) cannot be easily understood. Figure 5.1 presents Sartori's levels of abstraction (Sartori 1970) applied to the concept of governance.

Moving down the "ladder of abstraction" (Sartori 1970), many indicators have tended to approach the "real" meaning of governance focusing on sub-dimensions or specific sectors within which governance is supposed to operate. The Corruption Perception Index (CPI) for instance, is a mono-dimensional index of governance and it measures a specific aspect of governance, that is the level of corruption (Malito 2014). Indicators have an intrinsic prescriptive function that often obfuscates their descriptive function. Many indicator providers and institutions in fact agree that fighting corruption and promoting development is instrumental to improve "good governance". Yet, what is methodologically

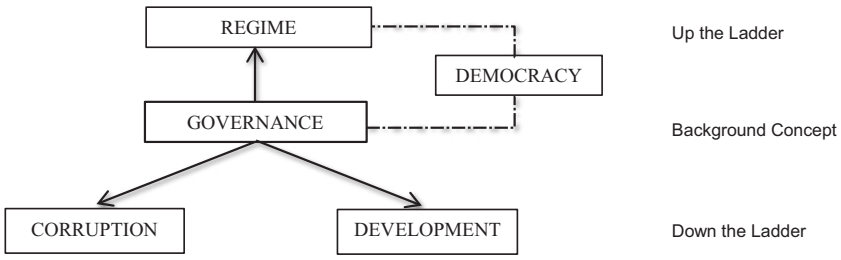

Fig. 5.1 Governance: up and down the generality

obscure is the operationalisation of governance through variables that are usually considered as explanatory or intervening variables within the semantic field. This misconception affects the clarity of the governance argumentation in a way that makes the relationship between governance and its sub-dimensions tautological. This critique is particularly important, as CPI holds a strong prescriptive function, but, its descriptive capacity remains problematic. It does not measure corruption, but rather the subjective perceptions of a coalition of experts and business elite. What is more, CPI measures different things, such as the frequency of corruption, the damage to private business caused by corruption, and both their frequency and value (see Malito 2014).

Moving up the ladder of abstraction (Sartori 1970), the concept of governance confronts the notion of democracy. While the UN defines governance as functional to implement democracy, some scholars claim, indeed, that democracy is a form of governance (Schmitter and Karl 1991). As result of this conceptual indeterminacy, indicators reproduce such a hybrid vision of democracy and its relationship with governance. Indices of governance usually incorporate measures of democracy (the Democracy Index is one of three measures used by the Bertelsmann Foundation to build the Sustainable Governance Indicators) and indices of democracy equally incorporate aggregate measures of governance. By comparing the categories,[6] of which some measures of governance and democracy are composed, the same underlining conceptualisations are used to capture the essence of both concepts.

Continuing the move up the abstraction ladder, the concept of governance confronts the notion of international order and political regime. Also, the relationship between these concepts is unclear. The most important characteristic of governance should be the existence of most divers, hierarchical, and non-hierarchical modes of political steering, as well as the non-centrality of the state. According to Rosenau, governance underlies a form of political order, a "system of rules with transnational consequences" (Rosenau 1995). Governance is also seen as a way of governmentalising international regimes (Zanotti 2005), as the manifestation of a broader agenda composed of sectorial international regimes (Gehring 1994). Within the discipline of International Relations (IR), governance has also been defined as a by-product of the demise of traditional agency and authoritative rules (Finkelstein 1995); there is no agreement on whether governance holds an intrinsic global connotation, or whether "global" should be considered only one of the potential "adjectives"[7] of governance (Rosenau 1995). Governance indicators, as a result, have established a contradictory relationship with the supranational characteristic of the concept. Few attempts have tried to differentiate between "global" and "national" governance and classical authoritative methods of

regulation. So, contrary to the globalisation discourse, an important trend of measuring aspects of governance is still embedded in national connotations, as pointed out by Cooley in Chap. 3 of this Handbook.

What Do Stateness Indicators Measures?

The measuring of governance has recently been complicated by the rise and proliferation of state fragility metrics. The term "stateness" refers to "the institutional centrality of the state" (Evans 1997), made of two basic conditions: the organisational capacity to formulate independent policies (Nettl 1968) and a coherent institutional framework indispensable to promote social cohesion.[8] The latter two conditions presuppose the classical attribute of decisive control over the application of authorised force within the territory. While theorising the state is an arduous enterprise challenged by conceptual and theoretical complexities, many scholars and practitioners have approached the problem of state weakness through the lens of empiricism. In less than two decades, the state failure discourse has produced, at least, four variants of "low stateness" (Evans 1997). As illustrated in Table 5.5, the notions of state fragility, weakness, failures, and collapse are used in parallel to frame the meaning of non-consolidated, low, or precarious stateness. As result, a variety of definitions and notions compose the matrix of this background concept.

The term "state fragility" describes a general incapacity to "deliver core functions to the majority of its people, including the poor"(Department for International Development, DFID 2005, p. 7). "'State weakness' refers to the lack of essential capabilities to implement 'critical' government responsibilities" (Rice and Patrick 2008, p. 8). It yet also encompasses the incubation of "ethnic, religious, linguistic, or other intercommunal tensions that have not yet, or not yet thoroughly, become overtly violent" (Rotberg 2003, p. 4). "State failure" instead is "a serious political crisis" (Goldstone et al. 2000) marked by the escalation of a military conflict. The state fails when the central authority loses its monopoly on the legitimate use of force. Rotberg described failed states as "tense, deeply conflicted, dangerous, and contested bitterly by warring factions" (Rotberg 2003, p. 5). While "state weakness" holds "a temporal or situational" connotation (Goldstone et al. 2000), "state failure" has a structural and enduring character. "State collapse", finally, represents an extreme case of state failure, marked by the complete erosion of public authority and legitimacy. A collapsed state differs from a failed state in its institutional character (Milliken and Krause 2002). When a state fails, the central authority is still existent, but extremely weak, whilst a state collapse reveals a "complete vacuum of authority" (Rotberg 2003, p. 5).

Table 5.5 Variants of low stateness: definitions

Variant of "low stateness"	Focus on	Definitions
Fragility	Core functions	"A State is fragile if the government cannot or will not deliver core functions to the majority of its people, including the poor" (DFID 2005, p. 7).
Weakness	Critical government functions	"Inherently weak because of geographical, physical, or fundamental economic constraints; basically strong, but temporarily or situationally weak because of internal antagonisms, management flaws, greed, despotism, or external attack; and a mixture of the two" (Rotberg 2003, p. 4).
Failure	The loss of the monopoly of violence	"Failed states are tense, deeply conflicted, dangerous, and contested bitterly by warring factions. In most failed states, government troops battle armed revolts led by one or more rivals. Occasionally, the official authorities in a failed state face two or more insurgencies, varieties of civil unrest, different degrees of communal discontent, and a plethora of dissent directed at the state and at groups within the state" (Rotberg 2003, p. 5).
Collapse	Implosion	"A collapsed state is a rare and extreme version of a failed state. Political goods are obtained through private or ad hoc means. Security is equated with the rule of the strong. A collapsed state exhibits a vacuum of authority. It is a mere geographical expression, a black hole into which a failed polity has fallen" (Rotberg 2003, p. 5).

Source: Amended version of Malito (2011, p. 61)

Measures of Stateness

The connection between governance and democracy that informed the first generation of governance measures lost its centrality when the process of democratisation in Africa and Middle East revealed elements of crisis (Chesterman et al. 2005). The production of governance indicators supported the positivist assumption of setting the benchmark and stages of a linear pattern of development. Yet, the limited achievements of the third wave of democratisation led many scholars and stakeholders to rethink the governance-democratisation nexus. State failure was considered a crisis of governance (Chesterman et al. 2005), which reinforced the search for good governance instruments and early-warning indicators of state weakness. However, whereas the category of the "failed state" emerged in reaction with the uncertainties posed by the process of democratisation, it

was only with the declaration of the Global War on Terror, that the state failure paradigm became increasingly dominant, both in the academic and public debate.

As a result of these shifts, a number of new measures have been produced on measuring state capacities. While only little empirical evidence supported the formulation of the fragility-security nexus, a number of research projects tried to create early-warning instruments capable of bridging the gap between theory and practice (Carment et al. 2009). Some of the most important measures of stateness are illustrated in Table 5.6.

One of the first aggregated measures of state weakness was inaugurated by the US's Central Intelligence Agency in the early 1990s and launched by the Centre for Global Policy at George Mason University. It was the Political Instability Task Force (until 1994 named State Failure Task Force). The Task

Table 5.6 Measures of stateness

Index	Provider	Aim	No. indicators	Coverage
Country Indicators for Foreign Policy Fragility Index	Carleton University	Assist donor community.	83	183
Low-Income Countries Under Stress	World Bank	To secure the bank's investment in developing countries affected by a set of government and governance endemic crises (Carment et al. 2009).	6	73
Fragile States Index (until 2013, Failed State Index)	Fund for Peace	"To have meaningful early warning, and effective policy responses to the emerging problem of state fragility" (The Fund for Peace 2014).	12	178
Index of State Weakness in the Developing World	Brooking Institute	Cognitive function: analyse the world's most vulnerable countries.	20	141
State fragility Index	George Mason University	USAID's request to better monitoring and managing the sources of state fragility.	14	162
Political Instability Task Force	George Mason University	Formulating early-warning instruments for the US Foreign Policy.	4 datasets, 1300 variables ca.	157

Source: Own compilation

Force was created in response to the need in formulating Early Warning Research for US Foreign Policy, especially in Africa. The task force annually produces four datasets on internal wars that are conceptualised as forms of both political instability and governance failure.

In 2002, the WB started to rank fragile states within the Low-Income Countries Under Stress (LICUS) programme to identify countries that perform badly on two core dimensions (low income and poor CPIA performances). The LICUS initiative was created to support WB decisions on investments in developing countries affected by endemic crises (Carment et al. 2009).

Rotberg provided one of the most important qualitative categorisation of state failure. In 2003, after five years of research on the Failed State Project at Harvard University, he defined the dividing lines between state collapse and general phenomena of state weakness. In 2004, the Fund for Peace launched the Failed States Index (renamed in 2014 into Fragile States Index), an annual index published by Foreign Policy, to establish a meaningful early-warning mechanism, and effective policy responses to the problems of state fragility (The Fund for Peace 2014). The index is based on 12 indicators, created through means of content analysis and calculated for 178 countries (edition 2014).

In 2007, the Center for Systemic Peace and the Center for Global Policy at George Mason University launched the State Fragility Index. In the same year, the Canadian International Development Agency commissioned the formulation of an Index of State Fragility by the Carleton University to support capacity-building within the Canadian donor community. The index assesses the fragility of the state according to three criteria: the authority, the legitimacy, and state capacity. The 2011 edition ranked 198 states using 83 indicators. In 2008, also the Brooking Institute published its first *Index of State Weakness in the Developing World* (Rice and Patrick 2008), a multi-dimensional index composed of 20 indicators, including 141 countries, and oriented to target especially the wider public.

The Semantic Field of Stateness

Analysing the relationship between state fragility and a set of associated phenomena, the boundaries of this semantic field raise important questions. The first problem is related to the lack of a systematised concept within which the variants of low stateness are located. The lack of systematisation is rooted in the inconsistent theoretical status of the literature. The notions of state

fragility and state weakness are rather "orphan" concepts, detached from vital contributions made by social and political analyses to the theory of the state. Hence, the literature has developed prototypes of state performance (Sanín 2011) that incorporate causes, symptoms (The Fund for Peace 2014), and consequences of fragility into the same metric. At the same time, also the relationship between variants of low stateness is subject to different interpretations, and different terms are used to identify the same phenomena. For instance, many providers use the term "fragility" to define the "fundamental failure of the state", without explaining why they use two different terms to represent, in the end, the same phenomenon (Batley and Mcloughlin 2010; DFID 2005). As a result of this differentiation and theoretical inconsistency, the debate has been trapped into subjective, judgemental (Gruffydd Jones 2013; Menkhaus 2010; Sanín 2011), and political (Barakat and Larson 2014) connotations.

Moving up the ladder of abstraction, the relationship between governance and stateness creates others problems. According to some scholars, the definition of good governance overlaps with the definition of state capacity (Hanson and Sigman 2013), and the same dimensions or categories (Economic, Political, Social, Security) are employed to operationalise both concepts. The *Global Governance Report* released by the Hertie School of Governance (2014), for instance, introduces the concept of "governance readiness", to identify the "resilience and adaptiveness" of private and public actors to create or maintain "the conditions in which problem-solving is possible as the resources of different state and non-state actors are brought together" (2014, p. 20). With the exception of security, the report equates the conditions for statehood with the "object" of governance. The four conditions that governance should fulfil are the provision of welfare, infrastructure, sustainability, and societal integration. Hence, the concept of governance remains fundamentally state- and authority-centred, focused on auxiliary and administrative functions of the classical state machinery and exercise of power. As an overall consequence of the lack of theoretical specification, the semantic field of stateness remains opaque; the internal conceptual logics flawed and its external definitional boundaries porous.

…How? Methodological Considerations

The current literature on rankings, performative indicators, and indices questions the overall quality of measurement activities (Galtung 2006), the selection of variables, the capacity of the selected indicators to capture the core

concepts, and even the reliability of aggregation methods employed. Key difficulties in measuring governance and stateness have risen. However, much of the methodological issues with governance and stateness indicators are rooted in the epistemic and conceptual challenge of operationalising these concepts in a valid, representative, and reliable way.

The early measures of governance (WGI, ICRG, CPIA, CPI) faced the conceptual problem of measuring phenomena intrinsically "immeasurable" (Bell and Morse 2008). As a consequence, indicator providers created a set of methodological shortcuts to get to the measurable core of governance.[9] This way of proceeding was accompanied by two fundamental flaws. First, the use of proxies in itself questioned the measures' capacity to capture the real meaning of governance. Second, the use of subjective, perception data raised greater criticisms. Subjective data were introduced into measuring governance in order to better frame the governance discourse and to solve the existent cognitive deficit. Governance emerged as an intangible concept, and considering the difficulty of attributing to this concept a concrete and countable denotation, subjective indicators (such as perception surveys) were extensively used. Scholars, however, doubted the capacity of perception-based indicators to correctly represent the phenomena under observation (Galtung 1998) without being themselves biased by the judgement of business elite or experts. Perception-based indicators have, moreover, been criticised for their failure to generate replicable data, indispensable for cross-country comparisons.

In response to these constraints of the first generation of indicators, the WB aimed for a second generation capable of creating more objective, disaggregated, nuanced, and contextualised solutions to the governance quests. According to Knack et al. (2003), this second generation was foreseen to fill the gap between measures and policy decision-making, and to make indicators more politically acceptable and replicable. The methodological and ontological choices of that second generation hence emphasise the need to create more objective measures, the combination of multiple data sources, and the composition of multi-dimensional indices.

As outlined above, in the 2000s, the proliferation of stateness indices intersected with the further advancement of such second generation. Following Mata and Ziaja (2009), two defining trends of that period need to be mentioned: first, maximalist definitions of state fragility included "ideas of good governance, democratic rule and extensive public service provision" and complicated "considerably the measurement of the phenomenon – [as] the more the state functions considered, the greater the variables and interdependencies to be controlled" (Mata and Ziaja 2009, p. 14). Second, minimalist

definitions oversimplified "the phenomenon and end[ed] up excluding elements that are crucial for validly representing a phenomenon" (Mata and Ziaja 2009, p. 14).

Moreover, many stateness indices aggregated existent indicators and data sources in order to operationalise new concepts. While the governance and development debate collected new data, the majority of indices of state fragility use the existent empirical data pool created by the first generation of indicators. As result, this third generation of metrics condensed the characteristic (Mata and Ziaja 2009) virtues and fallacies of the governance and development measures in a dangerous mix of ontological and conceptual incongruences. As a result of the continuous overlap between state and governance, and conflation into a common measure of different things, the descriptive capacity of these indicators of stateness is contested. For instance, measures of administrative capacity have been used to account for the fragility of institutions, without any differentiation between administrative infrastructure and state capacity. The Failed States Index, for example, uses both WGI and ICRG variables to measure the progressive deterioration of state functions, while the WB's LICUS initiative utilises CPIA ratings to determine the degree of state fragility.

…Why? Normative Demands and Policy Prescriptions

Measures of governance and stateness should be analysed in view of the normative and policy prescriptions they exert on the development agenda. They are both donor-serving concepts (Barakat and Larson 2014) supporting a normative model focussed on the promotion of a linear, global, model of liberal democracy. When this prescription became a policy imperative (Boutros-Ghali 1996), the UN attributed to democracy the instrumental function of fostering good governance. This process was supported by the implementation of traditional regulatory instruments (adhering to international regimes), and by the formulation of new policy instruments, including "benchmarking, co-regulation, voluntary codes of conduct and negotiated agreements" (Jordan et al. 2005).

The formulation of indicators capable of capturing the principles (inclusiveness, accountability, transparency, responsiveness, equity) or attributes (good, democratic, economic, corporate) of governance has been of paramount importance for the creation of the global governance model, as well for the reform of contemporary institutions and practices. Yet, despite the fact

that both indicators of governance and stateness support a normative model focussed on the value of democracy, two specific prescriptive orientations need to be considered.

Development Prescriptions

Since the emergence of the new donor enthusiasm of the 1990s, institutions and practitioners have used governance indicators to guide aid allocation, business investments, and donor priorities. Many governments of donor countries and risk rating agencies rely on the CPIA, originally created to evaluate the eligibility of developing countries for International Development Association (IDA) loans. The "Performance Based Allocation" (PBA), the system used to allocate IDA assistance, is calculated as a function of the country performance rating,[10] population and Gross National Income (GNI) (The World Bank 2007a). The country performance rating is built upon the aggregation of three indicators deriving from CPIA and the Annual Review of Portfolio Performance. The Millennium Challenge Account, indeed, relies on 17 indicators of governance quality divided into three categories (Ruling Justly, Encouraging Economic Freedom, and Investing in People). Of the 17 indicators, five derive from the WGI.

Yet, doubts have been raised about the capacity of this "one-best-way model of governance" (Andrews 2010) to represent what good governance means, and requires, for different countries. Both the legitimacy and effectiveness of performance indicators and political conditionalities have been questioned. Developing countries have been the recipients of aid and loan strategies to improve their standards of governance. Nevertheless, in some cases, the indicators employed to draft roadmaps to good governance lack the theoretical or empirical foundations to justify the recipes prescribed by the international organisations. In other cases, they are based on theoretical assumptions that have no legitimacy or relevance in developing countries (Andrews 2008). Throughout the 1990s, many influential practitioners, such as the World Bank emphasised the role of political conditionalities in creating incentives for good governance. However, the empirical evidence forced scholars and institutions to recognise that conditionalities could have a potential and predictable negative impact on the integrity of the recipient countries (Knack 2001), creating dependency syndromes, and eroding the overall quality of governance. The model of democratic governance adopted in developing countries during the 1990s, resulted in a "governance crisis" (Börzel et al. 2008). The reduction of the state to market-based mechanisms in conditions

of precarious administrative infrastructures inhibited the formulation of indispensable development plans. Poor economic infrastructure, low investments, and weak state architecture contributed to power distortions (such patrimonialism and predatory rule) in many transition countries. This crisis induced international institution and practitioners to formulate new development and aid strategies. As outlined above, within this reorientation, the formulation of stateness indicators has been pivotal to capture a set of new challenges.

Neo-trusteeship

Measures of stateness have been defined as instrument of neo-trusteeship (Caplan 2007; Fearon and Laitin 2004) and postmodern imperialism (Fearon and Laitin 2004). According to Laitin and Fearon (2004), the attribute of postmodern imperialism lies on the temporal horizon of the policy options for intervening in fragile state contexts. According to them, "the agents of neotrusteeship want to exit as quickly as possible, after intervening to reconstruct or reconfigure states so as to reduce threats arising from either state collapse or rogue regimes empowered by weapons of mass destruction" (Fearon and Laitin 2004, p. 7). The short-term and outcome-oriented horizon of policy interventions in the context of fragility holds a particular relevance to frame the normative prescriptions anchored to stateness metrics.

The problem of state weakness became a global security concern when the international community anchored the concept of international terrorism to the condition of state failure. The fight against terrorism and state failure became closely interrelated. The condition of state weakness became, hence, the condition justifying the new interventionary politics of the international community. If we consider the practice, and rhetoric, of some of the most important development agencies or institutions, there is no agreement on what state weakness implies. There has been, indeed, a wide global agreement in approaching state weakness as a security problem for the stability of democratic and effective states (Carment et al. 2008; Rice and Patrick 2008).

This security emphasis also emphasised the emphasis on intervention as state and social engineering machine, needing for outcome-oriented and short-term strategies of intervention. The US Agency for International Development (USAID), for instance, supports short-term solutions that mirror the strategic relevance fragile countries have for the US foreign policy (USAID 2005). While USAID recognises the need to deal with long-term policy prescriptions, its operational choices seem to encourage shorter planning horizons to support rapid and effective responses (USAID 2005). DFID

also focuses on short-term aims targeted at policy solutions for state fragility or the delivery of essential state functions. DFID developed a guide for working in fragile contexts that substantially redefines assistance criteria towards peace-building and state-building concerns, supporting security reforms, as well as the development of formal systems of rules and laws. Since 2005, also the OECD Forum on Aid Effectiveness acknowledges the relevance of state fragility for aid effectiveness. Finally, a number of international organisations and 138 countries signed the *Paris Declaration on Aid Effectiveness* that recognises the need to adapt aid strategies to post-reconstruction and capacity-building priorities.

In these contexts, indicators of stateness have been pivotal in capturing security risks and vulnerabilities of states in the context of "fragility". They guided the formulation of specific priorities and donor methodologies (Paris Declaration 2005). Yet, they also idealised a model of state strength as a guiding goal for international interveners. While classical forms of conditionality gradually lost importance, international practitioners promoted the use of more flexible, hybrid, outcome-centred aid instruments that "mostly focus on objectives, policy actions and standards, but leaves government more discretion over what the aid is spent on and how it is managed" (Manuel et al. 2012, p. 10). This soft mode, or positive conditionality, is built upon the respect of a set of criteria, like the harmonisation, local ownership, alignment, and mutual accountability of the process (Paris Declaration 2005). As stated by the Overseas Development Institute, the entry conditionalities established under the Paris Declaration "may relate to ethical standards (respect for human rights or democracy), public financial management, or the content of overall or sectoral strategy. They can also relate to bilateral objectives (supporting donor foreign policy positions)" (Manuel et al. 2012, p. 9). Addressing "ethical standards", foreign policy priorities, and political economy choices, the new conditionalities also allow international stakeholders to preserve a significant control over the political authority of the receiving country.

The Rule of Law Index, the Sustainable Governance Indicators, and the State Fragility Index

The present section analyses three distinct measures to describe some of the key characteristics of contemporary measures of governance and stateness. The measures under evaluation are The Rule of Law Index (World Justice Foundation), the Sustainability Governance Indicators (Bertelsmann Foundation), and the

Table 5.7 Comparison between RoLI, SFI and SGI

	State Fragility Index (SFI)	*Rule of Law Index* (RoLI)	*Sustainable Governance Indicators* (SGI)
Criteria	George Mason University	World Justice Project	Bertelsmann Foundation
Object	"A country's fragility is closely associated with its *state capacity* to manage conflict; make and implement public policy; and deliver essential services and its *systemic resilience* in maintaining system coherence, cohesion, and quality of life; responding effectively to challenges and crises, and continuing progressive development" (Marshall and Cole 2010, p. 7).	Rule of Law (proxy of governance, input): "the system of rules and rights that enables fair and functioning societies"(The World Justice Project 2012b).	Governance (input + output): the "government's capacity to deliver sustainable policies (executive capacity) as well as the participatory and oversight competencies of actors and institutions beyond the executive branch (executive accountability)" (Bertelsmann Foundation 2014, p. 14).
Purpose	Monitoring and managing sources of fragility	Cognitive function	Support OECD and EU sustainable policies
No. of variables	14	47	150
Coverage	162	99	41 (OECD and EU states)
Data	Objective	Subjective	Objective + Subjective
Method	Threshold standardisation	Min-Max value	Multilevel aggregation

Source: Own compilation

State Fragility Index (George Mason University). As summarised in Table 5.7, the three indices differ in a number of technical questions (coverage, aim, focus, numbers of variable, and methodology), but they also share important elements of continuity.

Demands and Purposes

The three indices respond to different demands and serve different purposes. The State Fragility Index (SFI) is a metric of political and economic instability that provides data about the level of instability for 162 countries (2011 edition). The development of the SFI was stimulated by USAID's request to

better monitor and manage issues of state fragility. The original initiative elaborated on the need to define the "prevailing perceptions of increasing global disorder" (Marshall and Cole 2008). The SFI forms part of the *Global Report on Conflict Governance and State Fragility* (Global Report), annually published by the Center for Global Policy at the George Mason University. The report claims that the Index is indispensable to define the challenges of the contemporary world order and to analyse the overall systemic performance of states in the "era of dynamic globalization" (Marshall and Cole 2008). The index calculates state fragility through 14 indicators grouped under four dimensions: economic development, security, governance, and social development. Each indicator, however, is evaluated on the basis of two criteria (effectiveness and legitimacy) essential to capture the level of state strength or fragility.

In 2010, the World Justice Project launched the Rule of Law Index (RoLI), an aggregate measure of the effective exercise of authority that can be considered as an approximation to the value of governance. The index is composed of 47 indicators grouped into nine clusters (Constraints on Government Powers, Absence of Corruption, Open Government, Fundamental Rights, Order and Security, Regulatory Enforcement, Civil Justice, Criminal Justice, Informal Justice), available for 99 countries. The key aim of the RoLI is to provide a comprehensible and accessible diagnostic tool, capable to strengthen the rule of law worldwide and within states. The RoLI looks at a nation's adherence to the rule of law from a societal perspective. This index holds a cognitive function, that is, the necessity to map the understanding and perception-based judgements of this concept. The Index, in fact, is not limited to the legal professional field, meaning that it does not exclusively target lawyers and judges. It also addresses the normative purpose of making "rule of law advancement as fundamental to the *thinking* and work of other professionals as it is to lawyers and judges" (The World Justice Project 2012a). The Index supports, indeed, the WJP's activity in sponsoring judicial reforms or access to the legal frameworks.

Since 2012, the Bertelsmann Foundation produces the Sustainable Governance Indicators (SGI), a composite measure based on 43 governance indicators grouped in three dimensions/indices (Policy Performance, Democracy, and Governance Indices). The indicators are available for the 41 OECD and EU countries. The 43 indicators used within the three indices are grouped in 13 clusters (Electoral process, Access to the information, Civil rights and political liberties, Rule of Law, Economic policies, Social Policies, Environmental Policies, Steering capability, Policy Implementation, Institutional Learning, Citizens, Legislature, Intermediary organisations). The SGI is a policy-oriented monitoring instrument with the specific aim to

assist the EU and OECD in formulating and implementing sustainable policies (Bertelsmann Foundation 2014). While the RoLI adheres to a normative priority, the SGI aims to configure precise policy solutions to governance issues. SGI provides in fact a cross-national comparative survey in order to explore how governments target sustainable development.

Concepts and Definitions

The three indices offer three different interpretations about what governance and stateness are, and about their interrelation.

First, the *Global Report* presents one of the most articulated and well-defined meanings of state fragility. According to Marshall and Cole (2009), in order to evaluate the integrity of states, scholars should consider the interconnections between governance, conflict, and development. The index also sponsors the need to enhance a systemic analysis of global trends, contrary to focussing "exclusively on individual or dyadic (bilateral) analysis, that is, on the conditions relevant to a particular country or state or relative to the interactions of two" (Marshall and Cole 2008, p. 3). The report, thus, sheds more light on the relationship between concepts and phenomena rather than formulating new definitions. Indirectly, it promotes a vision of democracy and autocracy as forms of governance, located at the opposite sides of a virtual governance spectrum. Governance, here, is considered one of the main criteria to calculate the fragility of the state. The SFI introduces also the important methodological innovation of considering state performance in terms of effectiveness and legitimacy. The index is also able to take into consideration how global trends affect state performance.

Second, the RoLI represents a mono-dimensional measure of governance. Both the World Bank and Mo Ibrahim Foundation have incorporated the rule of law into their respective measures of governance (respectively into the WGI and the Ibrahim Index of African Governance). The World Justice Project indeed has created a specific and innovative measure capable of capturing how the rule of law is experienced worldwide. It defines the rule of law as a system composed of four universal principles.[11] Whereas the "system of rules" implies that structural conditions are necessary to execute governance, the index systematises these four principles by exclusively using outcome indicators. It operationalises the index considering policy outcomes that the rule of law produces. These policy outcomes have been grouped in nine clusters.[12]

Third, for the Bertelsmann Foundation, governance refers to the "government's capacity to deliver sustainable policies (executive capacity) as well as the participatory and oversight competencies of actors and institutions beyond the executive branch (executive accountability)" (Bertelsmann Foundation 2014, p. 14). As a result, governance is understood as part of the broader pattern of sustainability, which brings together both input and outcome indicators. The SGI is built on three indices of Democracy, Policy Performance, and Governance.[13] While Democracy and Governance Indices use input and process measures of administrative capacity or democratic functioning, the Performance Index is based on 16 policy outcomes variables (in matter of economic, social, and environmental policies).

Data Gathering and the Quality of Data

Differences exist among the three examples also in view of the strategy adopted by the three providers to gather the necessary information and data. The RoLI is based on primary subjective sources, derived from two surveys (The General Pollution Poll and the Qualified Respondent's Questionnaire), produced by The World Justice Project. These two surveys collect a lot of information concerning the experiences and perceptions of citizens and legal experts with the rule of law.

The SGI uses both subjective and objective data. Hard data derive from Eurostat and the OECD. Yet, quantification is balanced by the formulation of country reports that synthesise the information provided by a pool of country experts responding to a specific questionnaire. The SFI employs only hard data. Each dimension of state fragility (security, political, social, and economic) presents both an effectiveness and legitimacy score that result from the aggregation of different data sources.[14]

Standardisation

In order to normalise data, the SFI uses a simple standardisation method, which establishes threshold values based on quintile cut-points. Nevertheless, from the document available online, it is not possible to appreciate the transparency of the process. The Global Report mentions that the fragility score has been normalised with a set of threshold values, formulated on the baseline year 2004. However, no other information is available to evaluate the consistency of this procedure. On the contrary, the Bertelsmann Foundation and World Justice Project offer complete information about their standardisation process and methodologies.

The RoLI relies on a simple normalisation, where the questions are first mapped by 47 indicators, and then codified by min-max values (0–1). Once normalised, the individual variables are integrated into sub-factors and factors, using simple averages. The same weight is assigned to questions (answers) formulated in the two datasets, the General Population Poll and the Qualified Respondents' Questionnaires.

The SGI presents a "multilevel aggregation" (Schraad-Tischler, and Seelkopf 2014) that combines both qualitative and quantitative information. In order to ensure the comparability of the data, the index standardises all the components through a linear function that transforms the value into a scale ranging from 1 (min) to 10 (max). To ensure comparability over time, thresholds are also calculated. The index is validated by adjusting the median (min and max value of the middle 50% of the distribution) by an amount equal to 1.5 times the interquartile range. The aggregation is made according to a scheme that aggregates data sources for each of the nine criteria (economic policies, social policies, environmental policies; electoral processes, access to information, civil rights and political liberties, rule of law; executive capacity, executive governance), attributing the same weight for both qualitative and quantitative data. In the same way, the score of each dimension, or index (the Policy Performance, Democracy and Governance), is produced.

Criticisms

Generally, the three indices face the common dilemma of measuring concepts that escape a clear conceptualisation, although they also contributed to filling parts of existent definitional and conceptual gaps. Yet, at the same time, they also contribute to complicating the ontological status of governance and stateness. The present section hence tests the validity (capacity of the indicator to represent the concept), reliability (the capacity to perform the required functions under different conditions and to returns the same results), and comparability (possibility to use the index for cross and within country comparisons) of these three measures.

Internal Validity

First, both SGI and RoLI face the common criticism of using subjective data. However, the SGI balances this potential subjectivity bias by using hard data derived from Eurostat and OECD datasets. For the RoLI, "*sensitive questions* may be perceived as threatening by government officials or by respondents. In

the first case, government officials of certain countries may censor or condition the administration of questions because they are perceived as challenges to the regime" (Botero and Ponce 2011, p. 23). This tendency produces a sort of bandwagon effect, where some opinion may obfuscate the sincerity of the answer provided. This distortion is particularly significant when practitioners have to provide answers on the effectiveness of government, or levels of corruption. However, the WJP introduced an innovative research method, based on the idea of capturing citizen perceptions about the rule of law. Contrary to this broader group, many of the governance indicators of the first generation are based on polls of experts or businessmen that ignored the perception of many other societal components.

However, the use of "soft" or "hard" data is not sufficient to assess the internal validity of the indices. The SFI, for instance, relies exclusively on objective data but the construction of the index relies on subjective assumptions, for two fundamental reasons. First, because state fragility is not a systematised concept, but remains entangled into judgemental (Gruffydd Jones 2013; Menkhaus 2010; Sanín 2011) and policy-oriented (Barakat and Larson 2014) connotations. Second, indicators of state fragility have created new categories according to specific policy meaning and interests. As noted by Ziaja (2012), the SFI does not provide a descriptive definition of state fragility, but rather an operational one, distinguishing between two criteria (effectiveness and legitimacy) and four fields of applications (and respective economic, security, social, and political indicators). This operational definition results from conceptual uncertainty, but it is influenced also by the difficulty of collecting specific data on fragility. The SFI, as many of the state capacity measures, aggregates measures on the basis of existing indicators of governance and development (Ziaja 2012).

Reliability

The three measures present different levels of reliability, that is, the ability of the index to perform its required functions under different conditions. In this context, both statistical and conceptual reliability should be considered. The conceptual reliability of governance and stateness measures is particularly important if we analyse the extent to which the operationalisation of concepts adhere to the reality of many developing countries. The Bertelsmann Foundation, for instance, does not provide evidence for the statistical reliability of the SGI. The index bypasses, however, the potential problem of a low level of "conceptual reliability", recognising that the challenge of "sustainable

governance" rather affects highly developed states of the OECD and the European Union. Therefore, the index is limited to the EU and OECD countries (41 states).

RoLI and SFI are indices of global scale facing the dilemma of whether the criteria used to construct the indicators reflect visions of governance and stateness that may have different meanings across countries. The SFI does not mention the statistical reliability of the index, but the conceptual understanding raises two issues. First, the notion of state fragility adopted tends to objectivise two theoretical categories—legitimacy, and effectiveness. These two categories have been formulated to capture a particular model of sovereignty, according to which the state is sovereign because it is capable of monopolising the legitimate use of violence. "Effectiveness" and "legitimacy" are categories elaborated on the Weberian ideal-types of modern states state. The seductive attempt to use such ideal-types as ahistorical and invariant universal categories of causal analyses is at odds with the very conceptual rationale of the Weberian methodology that attributes to ideal-types an interpretative function (Weber 1991). Second, the definition of state fragility also tends to parameterise the relationship between one ideal-type of low stateness (state fragility) and the empirical reality. But this conceptual operation does not define the relationship between the other ideal types (strong stateness, for instance) and the empirical data. As a result, only one typology of low stateness (state fragility) has been parameterised and transformed into a new phenomenon (the fragile state). It has then been isolated from the conceptual understanding and historical, philosophical connotation of its original category, the state.

From a methodological perspective, the WJP points out that the index holds a good level of reliability, because for seven of the nine dimensions, the statistical consistency measured with the Cronbach alpha's coefficient is at 0.90 (the threshold for an affordable aggregation is 0.7). Yet, if we question the definition of the Rule of Law provided by the World Justice Project, we could come to different conclusions about the conceptual reliability of the index. The WJP approaches the Rule of Law "in terms of the outcomes that the rule of law brings to societies – such as accountability, respect for fundamental rights, or access to justice" (The World Justice Project 2012b). This definition indirectly assumes the existence of an agreed-upon definition of the rule of law, or at least, that the law brings to different societies always the same kind of outcome. A deep disagreement exists between theorists and legal experts about the existence of such a universal connotation. The World Justice Project assumes that four universal principles derived from international standards can account for the rule of

law in any society. However, the operationalisation in terms of outcomes raises ontological incongruence. First, it ignores the "tensions" faced in the "overseas land" (Costa and Zolo 2007), where the concept of rule of law is a product of colonial importation. Second, it ignores how different forms of law (i.e. the case of Customary Law) are not quantifiable according to the same criteria used to evaluate the effectiveness in western societies. The WJP's definition indeed tends to standardise the outcomes that the rule of law brings in some countries (in the majority of them), without considering whether a universal principle is relevant or valid to capture the plurality of local information. As argued by Ringer, "The commitment to a conception of the rule of law in terms, roughly, of the number of courts per capita overlooks the importance of the interaction of state (i.e., 'official') courts with informal for a of dispute resolution, which are long-lived and vital parts of community life. Thus, 'local' and 'foreign' understandings of the rule of law will give rise to different descriptions and prescriptions for the problems" (Ringer 2014, p. 183).

Comparability

The quality of data employed also affects the comparability of the instruments, that is, the possibility to use the indices for cross- and within-country comparisons. While the three indices allow for comparisons across countries, the RoLI, for instance, does not allow for within-country comparisons because the scores do not capture absolute values.

The SGI allows for within-country comparisons. The methodology is supported by the implementation of the website's interactive interfaces through which data is displayed in a very intuitive manner. The SGI allows policymakers to derive lessons, especially because the metric is structured around three sub-categories (Policy Performance, Democracy Index, and Governance) that facilitate comparative analyses.

Comparison is essential to promote conformity since it activates "(international) learning processes while at the same time cast[e] a spotlight on vital reforms for decision-makers and the public" (Bertelsmann Foundation 2014, p. 3). More specifically, the annual reports, country studies, and sector-based analyses released by Bertelsmann Foundation identify "forward-looking" reforms and practices. For instance, the Greece Report released by the Foundation in 2014 shows how data drawn from the three indices have been used to monitor the Greek policy performances and reform capacity after the adhesion to the Memorandum of Understanding (Sotiropoulos et al. 2014).

The standardisation used for the SFI allows for comparisons over time, but not across countries. On the methodological side, many indicators are rescaled when new indicators are added, or new data for years became available. Moreover, scores of the matrix are not scaled according to a homogeneous numerical representation. While the majority of indicators are rated according to a four-point scale (0 = no fragility, 1 = low fragility, 2 = medium fragility, and 3 = high fragility), only one indicator (Economic Effectiveness) is rated on a five-point scale (they include also 4 = "extreme fragility"). While the total Legitimacy and Effectiveness score are numerically represented, the matrix displays only the colour icons (green = 0, yellow =1, orange = 2, red = 3, black = 4) for eight indicators. Other three indicators are rated on a nominal (Regime Type and Regional Effects) and alphanumerical scale (Net Oil Production or Consumption) that contributes to the erosion of both the comparability and intuitive understanding of this instrument.

Conclusion

While there has been a profusion of metrics, the concepts of governance and stateness remain vague and narrowly interpreted. The chapter claims that such "creative disorder" reflects the multi-level and multi-agents' nébuleuse of governing actors and structures informing the neoliberal process of governance. Therefore, this chapter emphasises the relevance of evaluating both the normative demands and conceptual validity of these measures. "The first question that should occupy potential users of any governance indicator is not the size of the margins of error, but whether the indicators are valid measurements of what they purport to measure" (Thomas 2010, p. 37). On the contrary, many of the contemporary analyses hide behind technical issues a number of ideological and conceptual problems. This analysis shows that governance indicators do not always satisfy the standards of conceptual clarity, while stateness indicators are not even theoretically defensible.

The second question to be addressed when analysing indicators is how policy demands and normative prescriptions inform the construction of indicators. The study of indicators of governance has been often bounded to the search of descriptive solutions. The chapter shows indeed that both indicators of stateness and governance are donor-serving concepts (Barakat and Larson 2014) indispensable to steer how states and institutions deal with political crises that challenge the stability of the world order. During the 1990s, the crisis of governing stimulated the emergence of the governance debate. Here, a minimalist vision of the state-sponsored policy reforms

focussed on expanding the role of the market in government functions. While the (good) governance agenda was about "re-structuring the polity and politics" (Börzel et al. 2008, p. 7), the state fragility discourse introduced the idea of re-structuring the state by favouring short-term and outcome-oriented strategies. The crisis of governance, in turn, forced the academic debate to rehabilitate the old state category, which has become the object of a new form of political conditionality.

In conclusion, the third question to be addressed concerns the use and implications of these indicators. There is broad disagreement and scepticism about the concrete capacity of these instruments to guide policymaking. Yet, indicators have the intrinsic function of building and shaping public discourse and knowledge at large. This implies also the transformation of Western categories of governance and liberal democracy into parameters of an imposed universal mode of governance. This outcome is not a mere "consolation prize"; it rather represents one of the most important materialisations of the soft form of power exercised by indicators within the governance context.

Notes

1. First published as EUI Working Paper, Malito, Debora Valentina. The Difficulty of Measuring Governance and Stateness. *Robert Schuman Centre for Advanced Studies Working Paper*, no. 38 (2015). http://cadmus.eui.eu/handle/1814/36356. The working paper was originally prepared for the Workshop *Global Governance by Indicators: Measuring Governance and Stateness* held at the European University Institute (13-14 November 2014). In this final form, the chapter benefited from many insights and discussions. I thank the workshop participants, Nehal Bhuta and Gaby Umbach for their helpful comments.
2. This is the case of the World Bank, which conceptually emphasises the administrative and structural infrastructure of governance, but in practice, use performance indicators, to operationalise the *Worldwide Governance Indicators*.
3. The CPIA is based on 16 variables grouped in four clusters: (a) economic management; (b) structural policies; (c) policies for social inclusion and equity; and (d) public sector management and institutions.
4. Voice and Accountability, Political Stability/Absence of Violence, Government Effectiveness, Regulatory Quality, Rule of Law, Control of Corruption.
5. Since 2007, the WB has yet also introduced new "Actionable Governance Indicators", based on mono-dimensional measures that, focusing on specific aspects each, are better designed to assist policymaking.

6. For instance, *the Democracy Index*, provided by Economist Intelligence Unit presents 60 indicators grouped into five pillars: electoral process and pluralism, civil liberties, functioning of government, political participation, and political culture. The Sustainable Governance Indicators compiled by Bertelsmann, aggregate three indices (Policy Performance, Governance, and Democracy).
7. The expression "with adjectives" is used here to recall the methodological concern advanced by Collier and Levitsky (1997), to differentiate between democracy and its sub-dimensions.
8. The Weberian monopoly on the legitimate use of force is the fundamental characteristic of the state. However, the term stateness has been used by political theorists like Evans (1997) and Nettl (1968) to indicate the institutional centrality of the state. As a result, the monopoly on the use of force is essential to define the state. But to define the institutionalisation of this monopoly, we have to consider other two categories—the organisational capacity and the institutional coherence.
9. For a wider discussion of the methodological issues raised by corruption indicators cfr. Malito 2014.
10. The Country performance rating is equal to (0.8 *CPIA + 0.2* PORT) * (GOV/3.5).
11. The four principles are: "(1) individuals and private entities are accountable under the law; (2) the laws are clear, publicized, stable, and just; are applied evenly; and protect fundamental rights, including the security of persons and property; (3) the process by which the laws are enacted, administered, and enforced is accessible, fair; (4) Justice is delivered timely by competent, ethical, and independent representatives and neutrals" (World Justice Project 2014, p. 4).
12. The constraints on government powers; absence of corruption; open government; fundamental rights; order and security; regulatory enforcement; civil justice; criminal justice; and informal justice (World Justice Project 2014, p. 4).
13. The Democracy Index is composed by 15 indicators grouped into four pillars (electoral process, access to information, civil rights and political liberties, Rule of Law). The Policy Performance Index is calculated on 16 variables grouped into three basic sectors (economic, social, and environmental policy). The Governance Index is composed by 12 indicators, grouped into two pillars (Executive Capacity and Accountability).
14. Major Episodes of Political Violence dataset, the Political Terror Scales, Polity IV Project, Leadership Duration; Elite Leadership Coups datasets, World Development Indicators, Human Development Report; Structure of Trade; US Energy Information Administration, US Census Bureau and US Political Instability Task Force, UNDP Human Development Report, Religious Fractionalization dataset (Marshall and Cole 2008).

References

Andrews, M. (2008). The Good Governance Agenda: Beyond Indicators Without Theory. *Oxford Development Studies, 36*(4), 379–407.

Andrews, M. (2010). Good Government Means Different Things in Different Countries. *Governance, 23*(1), 7–35.

Barakat, S., & Larson, A. (2014). Fragile States: A Donor-Serving Concept? Issues with Interpretations of Fragile Statehood in Afghanistan. *Journal of Intervention and Statebuilding, 8*(1), 21–41. doi:10.1080/17502977.2013.770263.

Barnett, M., & Duvall, R. (2005). *Power in Global Governance*. Cambridge: Cambridge University Press.

Batley, R., & Mcloughlin, C. (2010). Engagement with Non-State Service Providers in Fragile States: Reconciling State-Building and Service Delivery. *Development Policy Review, 28*(2), 131–154.

Bell, S., & Morse, S. (2008). *Sustainability Indicators: Measuring the Immeasurable?* London: Earthscan.

Bertelsmann Foundation. (2014). *Policy Performance and Governance Capacities in the OECD and EU*. Gütersloh: Bertelsemann Shiftung. Retrieved from http://www.sgi-network.org/docs/2014/basics/SGI2014_Overview.pdf

Bertelsmann Stiftung. (2005). *Bertelsmann Transformation Index 2006*. Gütersloh: Bertelsmann Stiftung. Retrieved from http://dev.bertelsmann-transformation-index.de/fileadmin/pdf/BTI_2006_Brosch_re_GB.pdf

Bhuta, N. (2012). Governmentalizing Sovereignty: Indexes of State Fragility and the Calculability of Political Order. In K. Davis, A. Fisher, B. Kingsbury, & S. E. Merry (Eds.), *Governance by Indicators: Global Power Through Quantification and Rankings* (pp. 132–164). Oxford: Oxford University Press.

Börzel, T. A. (2005). European Governance–Markt, Hierarchie oder Netzwerk. In G. F. Schuppert, I. Pernice, & U. Haltern (Eds.), *Europawissenschaft* (pp. 613–641). Baden-Baden: Nomos.

Börzel, T. A., Pamuk, Y., & Stahn, A. (2008). *Good Governance in the European Union* (Berliner Arbeitspapier Zur Europäischen Integration). Berlin: Freie Universität zu Berlin. Retrieved from http://edocs.fu-berlin.de/docs/servlets/MCRFileNodeServlet/FUDOCS_derivate_000000000875/2008-7_Boerzel_Pamuk_Stahn.pdf?hosts=local

Botero, J. C., & Ponce, A. (2011). *Measuring the Rule of Law* (The World Justice Project – Working Paper Series, WPS No. 001). Washington, DC: The World Justice Project. Retrieved from http://foldvary.net/econ200/measuring.pdf

Boutros-Ghali, B. (1996). *An Agenda for Democratization*. New York: United Nations, Department of Public Information. Retrieved from http://www.un.org/fr/events/democracyday/pdf/An_agenda_for_democratization.pdf

Bovaird, T., & Löffler, E. (2003). Evaluating the Quality of Public Governance: Indicators, Models and Methodologies. *International Review of Administrative Sciences, 69*(3), 313–328.

Caplan, R. (2007). From Collapsing States to Neo-Trusteeship: The Limits to Solving the Problem of "Precarious Statehood" in the 21st century. *Third World Quarterly, 28*(2), 231–244.

Carment, D., Samy, Y., & Prest, S. (2008). State Fragility and Implications for Aid Allocation: An Empirical Analysis. *Conflict Management and Peace Science, 25*(4), 349–373.

Carment, D., Prest, S., & Samy, Y. (2009). *Security, Development and the Fragile State: Bridging the Gap Between Theory and Policy*. London/New York: Routledge.

Chesterman, S., Ignatieff, M., & Thakur, R. C. (2005). *Making States Work: State Failure and the Crisis of Governance*. Tokyo: United Nations University Press.

Collier, D., & Levitsky, S. (1997). Democracy with Adjectives: Conceptual Innovation in Comparative Research. *World Politics, 49*(3), 430–451. doi:10.1353/wp.1997.0009.

Cooley, A. (2014). The Emerging Politics of International Rankings and Ratings: A Framework for Analysis. In A. Cooley, & J. Snyder (Eds.), *Ranking the World. Grading States as a Tool for Global Governance* (pp. xiii–xiv). Cambridge: Cambridge University Press. Retrieved from http://www.lawschool.cornell.edu/cornell-IL-IR/upload/Cooley-PoliticsofRankings.pdf

Costa, P., & Zolo, D., (Eds.). (2007). *The Rule of Law. History, Theory and Criticism*. Dordrecht: Springer. Retrieved from http://link.springer.com/content/pdf/10.1007/978-1-4020-5745-8.pdf

Cox, R. W. (1980). The Crisis of World Order and the Problem of International Organization in the 1980s. *International Journal, 35*(2), 370–395.

Cox, R. W. (1997). Democracy in Hard Times: Economic Globalization and the Limits to Liberal Democracy. In A. G. McGrew (Ed.), *The Transformation of Democracy: Globalization and Territorial Democracy* (Vol. 1, pp. 49–71). Cambridge: Polity Press.

Curtin, D., & Wessel, R. A., (Eds.). (2005). *Good Governance and the European Union: Reflections on Concepts, Institutions and Substance* (Ius Commune Europaeum Vol. 49). Antwerp: Intersentia.

Cutler, C. (2013). Legal Pluralism as the 'Common Sense' of Transnational Capitalism. *Oñati Socio-Legal Series, 3*(4), 719–740.

DFID. (2005). *Why We Need to Work More Effectively in Fragile States*. London: Department for International Development. Retrieved from https://www.jica.go.jp/cdstudy/library/pdf/20071101_11.pdf

European Commission. (2001, July 25). *European Governance: A White Paper* (Commission of the European Communities, doc 01/10, COM 2001/428). Brussels: Commission of the European Communities. Retrieved from http://europa.eu/rapid/press-release_DOC-01-10_en.htm

Evans, P. (1997). The Eclipse of the State? Reflections on Stateness in an Era of Globalization. *World Politics, 50*(01), 62–87.

Fearon, J. D., & Laitin, D. D. (2004). Neotrusteeship and the Problem of Weak States. *International Security, 28*(4), 5–43.

Fine, B. (2009). Development as Zombieconomics in the Age of Neoliberalism. *Third World Quarterly, 30*(5), 885–904.

Finkelstein, L. S. (1995). What is Global Governance? *Global Governance, 1*(3), 367–372.

Freedom House. (2014). *About Freedom House*. Retrieved August 29, 2014, from http://www.freedomhouse.org/about-us#.VACLEbySx9k

Galtung, F. (1998). Criteria for Sustainable Corruption Control. *The European Journal of Development Research, 10*(1), 105–128.

Galtung, F. (2006). Measuring the Immeasurable: Boundaries and Functions of (Macro) Corruption Indices. In C. J. G. Sampford, A. Shacklock, C. Connors, & F. Galtung (Eds.), *Measuring Corruption* (pp. 119–156). Burlington: Ashgate Publishing Company.

Gehring, T. (1994). *Dynamic International Regimes. Institutions for International Environmental Governance*. Frankfurt am Main: Peter Lang.

Global Integrity. (2014). *Global Integrity Index*. Retrieved August 29, 2014, from http://www.gaportal.org/global-indicators/global-integrity-index

Goldstone, J. A., Gurr, T. R., Harff, B., Levy, M. A., Marshall, M. G., Bates, R. H., ... & Unger, A. N. (2000). *State Failure Task Force Report: Phase III Findings*. McLean: Science Applications International Corporation, 30. Retrieved from https://www.hks.harvard.edu/fs/pnorris/Acrobat/stm103%20articles/StateFailureReport.pdf

Gruffydd Jones, B. (2013). "Good Governance" and "State Failure": Genealogies of Imperial Discourse. *Cambridge Review of International Affairs, 26*(1), 49–70. doi: 10.1080/09557571.2012.734785.

Halperin, M. H., & Lomasney, K. (1993). Toward a Global" Guarantee Clause". *Journal of Democracy, 4*(3), 60–69.

Hanson, J. K., & Sigman, R. (2013). *Leviathan's Latent Dimensions: Measuring State Capacity for Comparative Political Research* (Manuscript, Maxwell School of Citizenship and Public Affairs, 1–41). Syracuse: Syracuse University. Retrieved from http://faculty.maxwell.syr.edu/johanson/papers/hanson_sigman13.pdf

Jachtenfuchs, M., & Kohler-Koch, B. (2004). Governance and Institutional Development. In A. Wiener & T. Diez (Eds.), *European Integration Theory* (pp. 97–111). Oxford: Oxford University Press.

Jessop, B. (1997). Capitalism and its Future: Remarks on Regulation, Government and Governance. *Review of International Political Economy, 4*(3), 561–581.

Jordan, A., Wurzel, R. K., & Zito, A. (2005). The Rise of "New" Policy Instruments in Comparative Perspective: Has Governance Eclipsed Government? *Political Studies, 53*(3), 477–496.

Kaufmann, D., Kraay, A., & Mastruzzi, M. (2011). The Worldwide Governance Indicators: Methodology and Analytical Issues. *Hague Journal on the Rule of Law, 3*(2), 220–246.

Knack, S. (2001). Aid Dependence and the Quality of Governance: Cross-Country Empirical Tests. *Southern Economic Journal, 68*(2), 310–329. doi: 10.2307/1061596.

Knack, S., Kugler, M., & Manning, N. (2003). Second-Generation Governance Indicators. *International Review of Administrative Sciences, 69*(3), 345–364.

Kooiman, J. (2003). *Governing as Governance*. London: SAGE Publication.

Lake, D. A. (2009). Relational Authority and Legitimacy in International Relations. *American Behavioral Scientist, 53*(3), 331–353. doi:10.1177/0002764209338796.

Lambsdorff, J. G. (2007). *The Institutional Economics of Corruption and Reform*. Cambridge: Cambridge University Press.

Levi-Faur, D. (2012). *The Oxford Handbook of Governance*. Oxford: Oxford University Press.

Lynn, L. E. J. (2013). The Many Faces of Governance: Adaptation? Transformation? Both? Neither? In D. Levi-Faur (Ed.), *The Oxford Handbook of Governance* (pp. 49–64). Oxford: Oxford University Press.

Malito, D. V. (2011). Somalia and State-Building: State Capacity or State Autonomy? A Critical Review of How to Decolonize African Studies. *Somaliland Journal of Peace and Development, 1*(1), 56–74.

Malito, D. V. (2014). *Measuring Corruption Indicators and Indices* (Robert Schuman Centre for Advanced Studies Research Paper 2014/13). Florence: European University Institute. Retrieved from http://dx.doi.org/10.2139/ssrn.2393335

Manuel, M., McKechnie, A., King, M., Coppin, E., & Denney, L. (2012). *Innovative Aid Instruments and Flexible Financing: Providing Better Support to Fragile States*. London: Overseas Development Institute. Retrieved from http://kms2.isn.ethz.ch/serviceengine/Files/ESDP/156733/ipublicationdocument_singledocument/87989a7f-7426-460b-ade1-d5261f645c4c/en/7884.pdf

Marshall, M. G., & Cole, B. R. (2008). *Global Report 2008. Conflict, Governance and State Fragility*. Vienna: Center for Systemic Peace. Retrieved from http://www.systemicpeace.org/vlibrary/GlobalReport2008.pdf

Marshall, M. G., & Cole, B. R. (2009*). Global Report 2009: Conflict, Governance, and State Fragility*. Vienna: Center for Systemic Peace. Retrieved from http://www.systemicpeace.org/vlibrary/GlobalReport2009.pdf

Marshall, M. G., & Cole, B. R. (2010). *State Fragility Index and Matrix 2009*. Vienna: Center for Systemic Peace. Retrieved from www.systemicpeace.org

Mata, J. F., & Ziaja, S. (2009). *Users' Guide on Measuring Fragility*. Bonn: German Development Institute. Retrieved from http://dspace.cigilibrary.org/jspui/handle/123456789/27411

Menkhaus, K. J. (2010). State Fragility as a Wicked Problem. *Prism, 1*(2), 85–100.

Milliken, J., & Krause, K. (2002). State Failure, State Collapse, and State Reconstruction: Concepts, Lessons and Strategies. *Development and Change, 33*(5), 753–774.

Murphy, C. N. (2000). Global Governance: Poorly Done and Poorly Understood. *International Affairs, 76*(4), 789–803.

Nettl, J. P. (1968). The State as a Conceptual Variable. *World Politics, 20*(04), 559–592.

Nye, J. S. (1990). Soft Power. *Foreign Policy, 80*, 153–171.

OECD. (2014). *Glossary of Statistical Terms. Governance*. Retrieved March 11, 2017, from https://stats.oecd.org/glossary/detail.asp?ID=7236

Paris Declaration. (2005). *Paris Declaration on Aid Effectiveness: Ownership, Harmonisation, Alignment, Results and Mutual Accountability*. Paris: OECD. Retrieved from http://www.mfdr.org/sourcebook/2-1paris.pdf

Rice, S. E., & Patrick, S. (2008). *Index of State Weakness in the Developing World*. Global Economy and Development. Washington, DC: Brookings Institution. Retrieved from https://www.brookings.edu/research/index-of-state-weakness-in-the-developing-world/

Richards, D., & Smith, M. J. (2002). *Governance and Public Policy in the United Kingdom*. Oxford: Oxford University Press.

Ringer, T. (2014). Development, Reform, and the Rule of Law: Some Prescriptions for a Common Understanding of the "Rule of Law" and its Place in Development Theory and Practice. *Yale Human Rights and Development Journal, 10*(1), 5.

Rosamond, B. (2000). *Theories of European Integration*. Basingstoke: Palgrave Macmillan.

Rosenau, J. N. (1995). Governance in the Twenty-First Century. *Global Governance, 1*(1), 179–209.

Rosenau, J. N., & Czempiel, E.-O., (Eds.). (1992). *Governance Without Government: Order and Change in World Politics* (Vol. 20). Cambridge: Cambridge University Press. Retrieved from http://journals.cambridge.org/production/action/cjoGetFulltext?fulltextid=6238380

Rotberg, R. I. (2003). *State Failure and State Weakness in a Time of Terror*. Washington, DC: Brookings Institution Press.

Rotberg, R. I. (2014). Good Governance Means Performance and Results. *Governance, 27*(3), 511–518. Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/gove.12084/full.

Sanín, F. G. (2011). Evaluating State Performance: A Critical View of State Failure and Fragility Indexes. *European Journal of Development Research, 23*(1), 20–42.

Sartori, G. (1970). Concept Misformation in Comparative Politics. *The American Political Science Review, 64*(4), 1033–1053. doi:10.2307/1958356.

Schmitter, P. C., & Karl, T. L. (1991). What Democracy Is... and Is Not. *Journal of Democracy, 2*(3), 75–88.

Schraad-Tischler, D., & Seelkopf, L. (2014). *Concept and Methodology – Sustainable Governance Indicators 2014*. Gütersloh: Bertelsmann Stiftung. Retrieved from http://www.sgi-network.org/docs/2014/basics/SGI2014_Concept_and_Methodology.pdf

Soederberg, S. (2004). *The Politics of the New International Financial Architecture: Reimposing Neoliberal Domination in the Global South*. London: Zed Books.

Sotiropoulos, D. A., Featherstone, K., & Karadag, R. (2014). *2014 Greece Report*. Bertelsmann Stiftung. Retrieved from http://www.sgi-network.org/docs/2014/country/SGI2014_Greece.pdf

The 2013 Legatum Prosperity Index. (2014). Dubai: Legatum Institute. Retrieved August 29, 2014, from http://www.prosperity.com/download_file/view_inline/2834

The Fund for Peace. (2014). *The Methodology Behind the Index*. Washington, DC: The Fund for Peace. http://fsi.fundforpeace.org/methodology

The Hertie School of Governance. (2014). *The Governance Report 2014*. Oxford: Oxford University Press.

The World Bank. (2007a). *IDA's Performance Based Allocation Systems: Options for Simplifying the Formula and Reducing Volatility*. International Development Association Resource Mobilization (FRM). Washington, DC: World Bank. Retrieved from http://documents.worldbank.org/curated/en/956181468136800628/IDAs-performance-based-allocation-system-options-for-simplifying-the-formula-and-reducing-volatility

The World Bank. (2007b). *World Development Indicators*. Washington, DC: World Bank Publications.

The World Bank Group. (2011). *CPIA 2011 Criteria*. Washington, DC: The World Bank Group. Retrieved from http://siteresources.worldbank.org/IDA/Resources/73153-1181752621336/CPIAcriteria2011final.pdf

The World Justice Project. (2012a, May 9). *Engagement*. Retrieved July 17, 2014, from file://localhost/Users/bonis/Library/Application%20Support/Zotero/Profiles/p3g9sax9.default/zotero/storage/W3JXWSVC/programs-events.html

The World Justice Project. (2012b, May 9). *What is the Rule of Law?* Retrieved July 16, 2014, from http://worldjusticeproject.org/what-rule-law

The World Justice Project. (2014). *WJP Rule of Law Index 2014*. Retrieved March 11, 2017, from http://worldjusticeproject.org/sites/default/files/files/wjp_rule_of_law_index_2014_report.pdf

Thomas, M. A. (2010). What Do the Worldwide Governance Indicators Measure. *European Journal of Development Research, 22*(1), 31–54.

Umbach, G. (2007). *Intent and Reality of a New Mode of Governance: "Get together" or "Mind the Gap"? The Impact of the European Employment Strategy on the Europeanisation of National Employment Policy Co-ordination and Policies in the United Kingdom and Germany.* PhD Thesis, University of Cologne.

Umbach, G. (2009). *Intent and Reality of the European Employment Strategy: Europeanisation of National Employment Policies and Policy Making?* Baden Baden: Nomos.

United Nations. (2014). *Global Issues at the United Nations*. Retrieved March 11, 2017, from http://www.un.org/en/sections/issues-depth/global-issues-overview/

US Agency of International Development (USAID). (2005). *Fragile States strategy*. Washington, DC: USAID.

Weber, M. (1991). *From Max Weber: Essays in Sociology* (New ed.). London: Routledge.

Weiss, T. G. (2000). Governance, Good Governance and Global Governance: Conceptual and Actual Challenges. *Third World Quarterly, 21*(5), 795–814.

Weiss, T. G., & Wilkinson, R. (2013). *International Organization and Global Governance*. New York: Routledge.

Wilkes, J. (2004). Corporate Governance and Measuring Performance. *Measuring Business Excellence, 8*(4), 13–16.

Zanotti, L. (2005). Governmentalizing the Post–Cold War International Regime: The UN Debate on Democratization and Good Governance. *Alternatives: Global, Local, Political, 30*(4), 461–487.

Ziaja, S. (2012). What Do Fragility Indices Measure? Assessing Measurement Procedures and Statistical Proximity. *Zeitschrift Für Vergleichende Politikwissenschaft, 6*(S1), 39–64. doi:10.1007/s12286-012-0123-8.

6

Sustainability, Sustainability Assessment, and the Place of Fiscal Sustainability

Paul Burger

Introduction

Along with "too-big-to-fail" risks, the European financial crisis and many national and international regulation efforts, a debate on fiscal sustainability has developed in recent years focusing mainly on public debt. The debate encompasses issues such as risk management, public expenditure and widely practised company tax "optimisation" strategies. Although it explicitly uses the term "sustainability", the debate itself has so far been only weakly embedded within the general debate on sustainable development. Against this backdrop, the present chapter strives to build some bridges between sustainable development in general and the specific topic of fiscal sustainability. First, it sets out the space for conceptions of sustainability in today's scientific market. Second, it addresses some major issues in the field of sustainability assessment to consider their potential strengths and limits in evaluating developments. In the last section, the chapter addresses the place of fiscal sustainability on the basis of the results from the first two sections. The chapter argues that fiscal sustainability is not a goal *per se* (it is not an intrinsic good), but should rather be seen as related to the resilience of the societal system and as contributing two important instrumental goods within the realm of "what to sustain".

P. Burger (✉)
University of Basel, Basel, Switzerland

D.V. Malito et al. (eds.), *The Palgrave Handbook of Indicators in Global Governance*,
https://doi.org/10.1007/978-3-319-62707-6_6

The Problem: The Meaning of "Sustainability" in "Fiscal Sustainability"

The discourse on sustainable development[1] (SD) has become more and more agenda-setting and has spread out to many societal fields since 1992. Along with this process, the term "sustainable" has gained enormous popularity in both the public and scientific realms. Against the backdrop of the popularity of the term, it is not very surprising that it has also spilled over into the field of public finance, especially in the aftermath of the various severe financial crises in recent years. "Fiscal sustainability" has become a key word in public and scholarly discourse on public debt and financial risk. To give one example from the public sphere, when announcing new forthcoming debts for the years 2015 onwards, after a period of staying in the black, the finance minister of the Swiss canton Basel-Stadt assured that the government would do its best to not leave the pathway of "sustainable finance" (Government of Kanton Basel-Stadt, GKBS 2014).

As is typical of popular and ubiquitously used terms, however, it is very often far from clear what a person is claiming when she uses the term. A speaker may be referring to the idea of sustainability as characterised by the World Commission of Environment and Development (WCED) report (1987) (often referred to as the Brundtland Report) and at the Rio Conference in 1992 (UN 1992). However, she could also simply be using the term colloquially, with the ordinary language meaning of "long-lasting" or "having a substantial impact". Alternatively, she could use the term as a *façon de parler*, that is, with no specific meaning at all. Moreover, even if the speaker uses the term in the tradition of the WCED report, it will still be far from clear what exactly she has in mind because there are different scholarly and public conceptions of SD. For example, there are many different action fields,[2] such as greening economy, changing mobility patterns, climate change mitigation and adaptation, agriculture and food, or sustainable universities. In each of these fields "sustainable" may have different meanings that relate the overall idea of sustainability to the specific requirements of the field in question. Within these fields, numerous institutionalising processes can be observed, such as sustainability reporting according to Global Reporting Initiative Standards, sustainability offices in universities, sustainability strategies of countries, and national and international consumer labels like Forest Stewardship Council, the Marine Stewardship Council, and fair trade labels. Moreover, the global community has stepped forward by replacing the rather unspecific goals stated in the Rio documents of 1992 with a set of 17 much more specific goals known as the Sustainable Development Goals.

For the sake of argument, let us assume that talking in terms of "fiscal sustainability" is more than using a rhetorical metaphor. The overarching function of such a term is to qualify or to assess something—be it development or a specific state of affairs or a human action. When talking about "fiscal sustainability", the speaker makes two implicit claims. First, she expresses her concerns that fiscal policy or the fiscal situation ought to be in accordance with some sustainability criterion, but currently it is not. The assessment criterion is something like "the fiscal situation ought to satisfy requirements XYZ". Second, and as a consequence of its violating this standard, the situation has to be changed to meet the requirements of "sustainability". From a functional semantic perspective, "sustainable" is an evaluative term even in its ordinary language sense.

Although looking at the function of the term helps to understand the type of content we are dealing with, it does not provide its meaning. Against the backdrop of the current inflationary use of "sustainable" and the opacity regarding its meaning, it is helpful to first step back and focus on some clarifications regarding the term. In addition to providing transparency of what we are talking about, stepping back is important in two respects when analysing the particular field of fiscal sustainability. First, looking at the options for understanding sustainability can help pave the way to capturing the relationship between the general topic of sustainable development and the particular subtopic of fiscal sustainability—if there is in fact such a relationship. Second, looking at these options can also help to better understand in what respect "sustainable" is an evaluative term and what type of evaluation or assessment we can expect.

The argument of this chapter unfolds in three steps. I will first sketch the most important basic features characterising societal discourse on sustainable development in general as well as the current most influential conceptions of sustainability in order to provide an outline of the options for a discussion on fiscal sustainability (section "Sustainable Development and the Range of Conceptions of Scientific Sustainability"). In a second step, I will address some methodological elements of sustainability assessment to demonstrate that these assessments are normally "if-then" statements and that the related quantitative indicators ("numbers") should be treated carefully and viewed from a critical distance (section "Sustainability Assessment"). Finally, I will argue for an understanding of fiscal sustainability as a subtopic within the overall term "sustainability". It stands for an instrumental good, not an intrinsic one. As such, fiscal sustainability is part of the broader issue of ensuring the state's action and reaction potential and ensuring the functioning of the system at the interface between intra- and intergenerational justice (section "Fiscal Sustainability").

Sustainable Development and the Range of Conceptions of Scientific Sustainability

Although the various elements making up the idea of sustainable development have roots reaching further back than the famous WCED report of 1987,[3] together with the documents from the Rio conference of 1992 this report strongly framed modern societal discourse on SD. Given the enduring global inequality and the resulting obligation to alleviate poverty, they both plead for huge development efforts, including economic development. There is, however, a substantial problem to tackle when taking this route. As traditional economic development is strongly based on natural resource extraction and produces many emissions and waste, and as natural resources and the sink capacities of ecosystems are limited, we face a development dilemma: if we use global natural resources for our economic and social development and continue to emit the way we have done so far, we may well be successful in developing global society towards increased wellbeing for *our* generation, but at the same time strongly undermine the capacity for future generations to do so as well. The so-called Western lifestyle (consumption patterns), as the claim goes, is not globally generalisable given the existing environmental resources and the inherent fragility of ecosystems. The remedy for avoiding such a development dilemma is a sort of paradigm change from "traditional economic growth" to "sustainable development" (or sustainable growth), which considers the legitimate claims of future generations to be able to live a good life too.[4]

The plea for sustainable development is based on a somewhat general risk assessment: there is a strong probability that severe scarcities will emerge or vulnerable ecosystems will substantially change their equilibrium states. If this happens, it will lead to harmful effects for human beings. Not only could the magnitude of harm potentially surpass the benefits today, but their occurrence would also result in strong inequality across generations: benefits for us today; harm for future generations. Hence, concerns regarding future (not-yet-alive) generations are at the basis of the rationale for public and scientific discussion of SD. Among other things, sustainable development consists in decoupling not only economic growth but also human wellbeing from resource and energy consumption and the related emissions.

This Brundtland and Rio-based idea of sustainable development is, however, a very abstract model. Although it claims to steer political and economic decisions, it primarily constitutes a general maxim about taking account of the long-term effects on future generations related to the scarcity of resources

and the fragility of systems. Armin Grunwald characterises this as a pre-deliberative role model (Grunwald 2009), that is, as a maxim that people can agree upon as a general framework condition for their deliberation (see also Christen 2013). Content, however, has to be filled in through deliberation. A second also quite generally stated maxim related to this general sustainability idea is directed at a more integrative way of policy making: to avoid future negative impacts, the current dominant sector-policy system should be given up in favour of integrated policies looking at the dynamic interfaces between social, economic, and environmental subsystems. These core general-maxim elements of SD obviously need to be further operationalised to really steer human decisions. This cannot be done without further conceptual elements.

Notwithstanding its abstractness, the content of the role model does not invite arbitrariness. The following six properties characterise it and the related societal discourse: (a) "sustainable" somehow qualifies development, that is, it is about shaping the future; (b) it is about ensuring human wellbeing for today's and future generations (intra- and intergenerational justice); (c) it includes considering risks stemming from the scarcity of resources and the fragility of systems[5] in its evaluative sphere; (d) it requires taking into account human-nature relations in our decisions; (e) "sustainable" is qualified as acknowledging uncertainty as a frame condition for policies and strategies (as a consequence of a–d); and (f) it calls for "change", that is, for societal transformation (see Burger and Christen 2011; Christen and Schmidt 2011).

Against the backdrop of these six general features, one might be tempted to distil three constitutive domains or aspects from the general idea of sustainable development. There is an overall goal of ensuring human wellbeing, and related to this there are concerns of intra- and intergenerational justice (a, b, c); there is additionally a scarcity of resources and a fragility of systems, and related to this there are concerns about vulnerability and resilience (c, d); finally, there is a call for societal transformation, and in particular for governance of SD (e and f). Most importantly, the overall idea of SD encompasses not only one or two of these components but its essence is built of all three.

Looking at science against this backdrop immediately reveals an extremely heterogeneous picture of how these different components have been conceptualised during the past 25 years.[6] There are those who overtly reject going along with the Brundtland tradition and its commitment to justice. Ehrenfeld (2004), for example, restricts the meaning of sustainability to nature's flows of energy and matter as blueprints for industrial production. On the contrary, Pearce and Atkinson (1998) defend the meaningfulness of SD as a "concept of equity between generations" (Pearce and Atkinson 1998, p. 7) by pointing out that it (a) considers longer time horizons than just overlapping

generations by searching for "*guarantees* that development will be sustainable" (Pearce and Atkinson 1998, p. 7) and (b) asks us to clarify what "should constitute the bequest between generations" (Pearce and Atkinson 1998, p. 8). Dobson (1996) structures the conceptual offerings into four categories: what should be sustained; why it should be sustained; the object(s) of concern (the scope of whose entitlements should be included, that is, only humans or also animals, etc.); and rules concerning substitutability between different types of capital. Whereas Dobson's *what to sustain* is goal oriented (what are the goods to be sustained over time?), Pearce and Atkinson argue for a capital-, hence means-, oriented approach[7] to SD by claiming that we should approach the preconditions for realising wellbeing rather than wellbeing itself (which for economists is mainly a utility function). Barry (1999) emphasises the relevance of population growth and understanding sustainability as a function relating population to resources. Kates et al.'s (2005) review distinguishes between approaches that (a) give *definitions* of a possible set of criteria, (b) frame SD as sets of *goals* ("what to sustain?"), (c) offer schemes for *measuring* it ("indicator systems"), or (d) interpret SD as a *societal practice* or even a societal movement. They conclude that the "concrete challenges of sustainable development are at least as heterogeneous and complex as the diversity of human societies and natural ecosystems around the world" (Kates et al. 2005, p. 20). In addition, there is also a strong stream of literature that focuses on the procedural side of transitions towards SD. According to this literature, SD is a paradigm case for a new type of societal steering (Lafferty 2004a; Voss et al. 2007), expressing a shift from traditional governmental steering to new modes of governance (Weale 2011) such as strengthening public participation (Meadowcroft 2004) or community-based decisions (Norton 2005). To add even more complexity to the different scientific approaches to SD, there is also a literature proposing understanding SD in terms of resilience and vulnerability (Gallopin 2006; Smith and Sterling 2008). According to this line of reasoning, SD is about safeguarding basic life-supporting system functions, be they of the global ecosystem or of societal systems (Robèrt et al. 2002).[8]

This brief journey through the scientific approaches to SD seems to not only paint a confusing and heterogeneous picture, but also to add evidence for what critics often find faulty with SD, namely that it is an empty concept inviting one to include anything that seems appropriate.[9] However, this picture does not seem to be accurate. Certainly, there is diversity but by no means arbitrariness. When we consider the three constitutive domains of SD identified above, we can rationally structure this multitude of approaches along these three lines, and by doing so develop the following overview.

A first group encompasses those approaches that explicitly base their conception of SD on theoretical elements regarding wellbeing and intra- and intergenerational justice. Wellbeing is the relevant intrinsic good, that is, the core brick of what to sustain, regardless of how it is further operationalised. Some even expand the scope of wellbeing beyond the domain of humans to also include sentient animals, as is done, for example, by Ott and Döring (2008). In any case, SD is conceptualised on the basis of theoretical elements taken from ethics. As there are a number of different conceptions of justice (e.g. distinguishing between the underlying metric of wellbeing and the principle of distribution) we will get different conceptions of SD. This group includes, among others, economics-driven approaches like non-declining welfare (Gowdy 2005) or non-declining capital stock (Pearce and Atkinson 1998) as well as the so-called minimal necessary condition approaches based either on basic needs (Meyer 2015)[10] or on capabilities (Burger and Christen 2011; see also Rauschmayer et al. 2012). Hence, it is possible to explain the obvious diversity as resulting from different rationales regarding justice.[11]

The second group takes its rationale from the existing substantial risks, especially those related to the fragility of systems. Climate change, degradation of ecosystems, and loss of biodiversity have become much more accentuated since the 1980s. Concerns regarding the carrying capacity and hence the resilience of the earth's ecosystem together with requirements to adapt to the already changing environment have become a focal point within the sustainability discourse. In addition, the vulnerability or resilience of social systems has gained attention (Berkes et al. 2006). There is also a debate on so-called state failure (see Call 2011),[12] but vulnerability has already been present for quite a while, for example within the Sustainable Livelihood Approach (Chambers and Conway 1992). The theoretical elements in this second group are normally based on some variation of system theory together with criteria regarding carrying capacities. According to this line of reasoning, safeguarding the life-supporting functions of natural and social systems (preventing major system breakdowns) is the relevant good, that is, the core brick of what to sustain—without denying the relevance of wellbeing (see, e.g. Robèrt et al. 2002). However, as neither resilience nor vulnerability are precisely defined terms (Gallopin 2006; Adger 2006; Janssen and Ostrom 2006) and as there are different approaches available for conceptualising human-nature systems (Ostrom 2009; Fischer-Kowalski and Haberl 2007),[13] we find strong variations within this group too. Nevertheless, we can again trace back the rationale for these variations.

The third group of sustainability conceptions takes its rationale from the component of "development" within SD. Development implies the existence

of societal processes and it requires building and transforming institutions, regulations, knowledge, skills, and the like. Sustainable development will not just happen, but requires steering efforts, regardless of how "steering" is further operationalised. The WCED and the Rio documents, for example, put strong emphasis on local participation and on overcoming the dominance of sector policies in favour of integrated policies. The literature in this third group focuses on favourable enabling conditions for such societal transformations. Following Lafferty's (2004a, b) claim of a correlation between form and function, the literature centres on topics such as adaptive (Armitage et al. 2009) or transition management (Kemp et al. 2007a), reflexive governance (Voss et al. 2006), and societal learning processes (Pahl-Wostl 2009), especially including participation (Meadowcroft 2004). These are goal-directed insofar as they are committed to the societal goals expressed by the idea of sustainability, but this literature does not contribute to the scientific debate on goals as such, but rather to the scientific debate on realising such societal goals. In line with Jürgen Habermas (2007), we can label these authors "SD proceduralists". Their approaches focus on societal capacities enabling change, especially looking at favourable conditions for deliberation and negotiation. Differentiation can be tracked down to different aspects and domains as well as to different theoretical approaches (e.g. institutionalism or governance). From the perspective of the three constitutive features identified above, we can understand their scientific contributions in terms of analysing relevant conditions and requirements for current and future societal capacity of action and reaction for change and adaptation.

Following the line of reasoning sketched above, we obtain a highly differentiated but nevertheless clearly structured and transparent picture of current sustainability approaches (Fig. 6.1).

This is not the place to further discuss their strengths and weaknesses. For the purpose of this chapter it suffices to say that I have argued against arbitrariness by revealing the different rationales behind the scientific variations, and especially (a) that the types respond to the three constitutive bricks within the topic of sustainability, and (b) that variations within these types come from the different emphases of established theories.

Sustainability Assessment

When talking in terms of "fiscal sustainability" and how it is related to the broader societal discourse on SD, links have to be drawn between public finances and (a) ensuring human wellbeing (concerns about intra- and

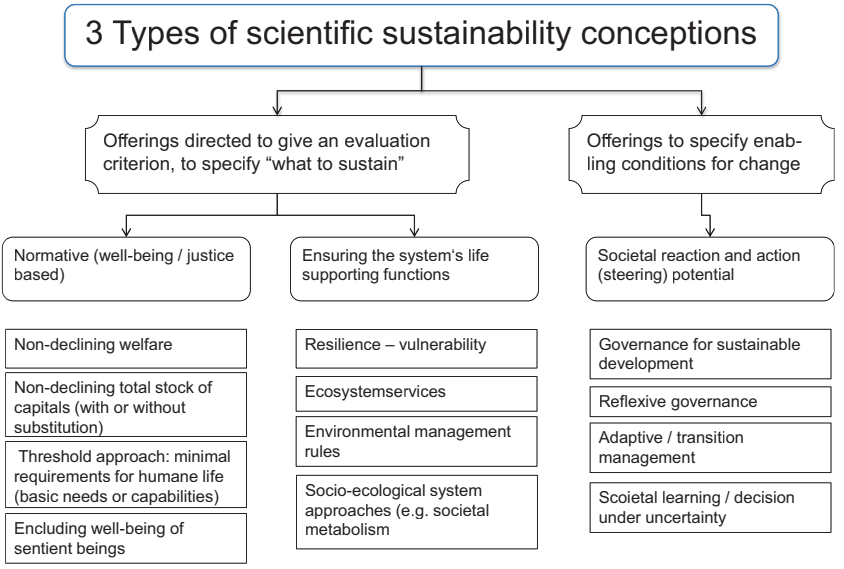

Fig. 6.1 The three types of sustainability conceptions and their internal differentiation. Source: Own elaboration

intergenerational justice), (b) concerns regarding vulnerability and resilience, and (c) societal transformation issues. Before looking at such possible links, I will first further highlight the type of information we gain from carrying out a sustainability assessment.

There is no dispute over the function of such an assessment: it should inform decision-making. A government wants to know the facts and figures which reveal whether the country's development points towards the goals as stated in its sustainability strategy. Actors would like to have a tool at hand that enables them to assess *ex ante* the possible consequences of their actions. In what follows the analysis will take this function for granted. However, a much less trivial issue consists in clarifying what "informing decision-making" means. Can we expect to provide arguments for best or second-best solutions on the basis of sustainability assessments?

Let us first have a look at three famous examples representing the challenges behind sustainability assessments. The first is nuclear power. It is easy to find two completely contradictory sustainability assessments regarding nuclear power: one supporting its sustainability; the other denying it (OECD 2000; Sustainable Development Commission UK, SDC-UK 2006). The Ecological Footprint (EF) is another case. The basic idea behind this is to translate energy and resource consumption into spaces needed for reproducing the amount demanded on a renewable basis and comparing

that space with what is available on our planet or within the country in question. Applying this scheme reveals that global human society needs "x" times the reproductive capacity of our planet and a specific nation "y" times its capacity. Following this approach to sustainability assessment, some Arab countries are said to be the most unsustainable and some countries of the South, like Gabon or Guyana, are said to be the most sustainable (Global Footprint Network, GFN 2010). However, from a wellbeing perspective, this amounts to saying that countries with a seriously deprived population are the most sustainable ones. The third case is the World Bank's (WB) sustainability criterion. According to this approach to sustainability assessment, China and Singapore are said to be the most sustainable countries. The ranking is based on a variation of the genuine savings approach (adjusted net savings—a version of weak sustainability) that measures the rate of (capital) savings for future generations in relation to investment in human capital, the depletion of natural resources, and damages caused by emissions. Whereas the EF leads to very poor countries being the most sustainable, the WB has dictatorships as its flagships.

The contradictory claims regarding nuclear power do not result from different understandings of what a nuclear power system is, that is, of its technical properties. They also do not involve different grades of risk aversion. They are the results of different criteria for "sustainability" together with different indicators used to operationalise the underlying criteria.[14] The rankings offered by the EF and WB are also fully dependent on their respective underlying assumptions. EF and WB primarily look at environmental resources, albeit from totally different perspectives. Democracy or societal transformation capacities are not part of either of these assessments.

Obviously, the result of a sustainability assessment is dependent on the underlying criteria and an evaluation will display how well or poorly the evaluative criteria employed are met. Sustainability assessments always provide "if-then" statements as information for decision-making: given that we have criteria set A to evaluate "sustainability", we get the result R. From section "Sustainable Development and the Range of Conceptions of Scientific Sustainability" above, we already know that we have to deal with different evaluation schemes for SD. Hopton et al. (2010) provide a good example from the field of environmental sustainability. They compare four assessment schemes (among them, the EF) and conclude that each assessment scheme has its merits and weaknesses in capturing the ecological domain. More importantly, the authors demonstrate that these assessment schemes pick up different aspects of the overall complex ecosystem. They also argue that we need all of them as they are mutually complementary.

Hence, we have good reasons to expect plurality within the field of sustainability assessment. As we have different options for sustainability conceptions and as any serious sustainability assessment will be based on a sustainability conception, we have to expect different assessment results. However, plurality is not to be confused with "anything goes". Plurality instead results from complexity itself together with different perspectives on it. If we require a sustainability assessment to include the three constitutive fields sketched above and if we agree in saying that there are different well-established ethical approaches but there is no grand unified theory of human-nature systems, then the consequence will be that we have differing results from sustainability assessments stemming from different assumptions.

From a decision-making point of view, however, such a "there are different assessment schemes with different results" situation could be seen as uncomfortable. There would then be no clear basis on which to build a decision. This is why aggregated indicators like GDP (as a measure of non-declining welfare) or the EF or WB are so attractive. They give a number that seems to allow benchmarking across countries. Such aggregated indicators meet the rational requirement that decisions should be based on a clear ranking of the options (X is better than Y and Y is better than Z). Only aggregated indicators can offer a clear ordering for decisions whereas a set of non-weighted indicators cannot. Even the strongest sceptics of the rational choice assumption behind this argument can concede the strength of such aggregated indicators.

However, first, plurality cannot be overcome in this way because it is still the case that each outcome of the assessments is an "if-then" statement which is dependent on the assumptions made. GDP, EF, and WB only capture specific factors, not the full range of relevant factors according to section "The Problem: The Meaning of 'Sustainability' in 'Fiscal Sustainability'". Second, there is a hidden assumption that we *can* always rank options in a better-worse order. Assume, however, that there are conflicting values, for example between claims for public participation and for the efficient functioning of a state. Suppose further that there are conflicts between, for example, economic efficiency and system resilience (a relevant issue in the discourse on energy transition) and between further improving wellbeing and resource scarcity. It is certainly logically possible to try to order all these conflicting goods, but it is also reasonable to argue that societal decisions are based on *balancing* these goods rather than finding best and second-best solutions because the complexity in question cannot be reduced to one decision-informing scheme.[15]

Further methodological elements for sustainability assessments are stated in the so-called Bellagio principles for sustainability assessment (Bellagio Principles, 1997). These include, among others, the requirement of a transparent

approach to the system in question (often including human-nature interactions) and that of a clear rationale for choosing indicator system "A" rather than "B". More generally, the following formal requirements for a sustainability assessment system can be postulated based on the Bellagio Principles: (a) they should strive to inform action; (b) they should be potentially adaptable to future developments; (c) there must be an underlying socio-ecological theory to capture human-nature systems; (d) there should be a systematic approach to representing the issue in question; (e) the normative principle(s) used should be made transparent; (f) there should be rules capturing the ecological limits; (g) a goal system for explicit assessment is needed; and (h) rules for "final evaluation", that is, aggregation rules or how to deal with multiple dimensions, should be stated (capturing what has been discussed above, at the very least). These formal requirements represent the characteristics of sustainability elaborated above as follows: criteria (c), (d) and (f) deal with scarcity and fragility embedded in human-nature systems; (e) and (g) address the inherent normativity of sustainability (wellbeing, justice); (a) and (h) are linked to "transforming societies", and (b) captures uncertainty and openness of development.

A more recent study on approaches to sustainability assessment of energy systems reveals that these formal requirements are only partially followed (Moser, 2014). The sample includes 30 papers/agency publications on sustainability assessments of the energy system and provides the following frequencies (Table 6.1):

Whereas (a) is quite well served, the crucial fields (e), (f), and (g), which capture core components of sustainability assessment and in particular its goal orientation, are covered to a much lesser degree. Moreover, none of them explicitly look at the energy system as a coupled human-nature system, which is astonishing. However, one could argue that all assessment approaches implicitly apply (c) so that no explicit argument is needed. The most important lesson to learn, however, is that we have not only a diversity of theoretical elements, as discussed so far, but we also have diversity based on how the methodological components of such an assessment are served or are taken into account.

Table 6.1 Frequency of SD Methodological Bricks in Sustainable Energy System Assessments

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
21/30	8/30	0	16/30	11/30	13/30	10/30	n.a.[a]

Source: Own compilation based on Moser (2014)

[a]No general statement on frequency can be made here due to there being too many different solutions and approaches within the sample

Again, the lesson to be learned is neither that "anything goes" nor that sustainability assessment is a worthless endeavour. Instead, the lesson is that the perspective provided by Hopton et al. (2010) should be more generalised. To put it simply, distrust any one-dimensional sustainability assessment. There may be cases in which clear and transparent rankings covering all the relevant possible options are achievable, so that decisions can be unambiguously informed. However, this is only possible when other relevant issues are deliberately left out, as our short discussion on GDP, EF, and WB clearly demonstrates. Complexity has to be reduced, but the question is to what extent and what information is lost in the process. There are good reasons to assume that relevant intrinsic and instrumental goods within the three domains of wellbeing, resilience of systems, and capacity for societal development are not independent—especially when we try to maximise any one of them.

An excursion to a thought experiment presented by Colin Farrelly (2007) could add additional force to this line of reasoning. According to John Rawls, the most relevant primary good to maximise is public participation. The more people can participate in voting and deciding, the better off they are (given the background assumption that they are able to pursue other aims as well). Against the backdrop of this premise, Farrelly invites us to look at California as a case in point. California is geographically large, has a big population, and its democratic activities bind financial and other resources. We cannot maximise participation without bringing more financial means into the game. But maximisation of participation does not generate financial resources. Hence, efforts to maximise democratic practices could lead decision-makers to reallocate resources from other fields, for example social programmes. This could then potentially have negative effects on the entire population, including those who are deemed to be the worst off or the poorer part of the population, thereby undermining the premise (that they are able to pursue other aims as well) on which the efforts for maximisation is built. The idea of SD points precisely to this problem: attempts to maximise wellbeing have to take into account the availability of resources and restrictions arising from efficiency considerations, for example. There are multiple components that need to be balanced and the reduction of complexity suggested must pay attention to the possibility of trade-offs emerging.

To sum up, we can say that sustainability assessments are "if-then" statements in at least three respects. They are dependent on normative elements regarding their evaluation criteria, they are dependent on elements regarding the representation of the system, and, since we are dealing with the future, they are always accompanied by uncertainty. It is reasonable to look of SD as requiring trade-offs rather than best or second-best solutions. Accordingly,

sustainability assessments can build a systematic information basis for societal deliberation and negotiation but they can hardly inform decision-making directly by delivering best or second-best solutions.

Fiscal Sustainability

There are three possible meanings of "sustainable" in "fiscal sustainability". First, its use may simply be due to the popularity of the term and it brings with it no specific content. Second, "fiscal sustainability" can be a specific discourse related to a long tradition of financial risk assessment without any clear relation to the overall sustainability discourse. Here, the meaning of sustainability is strongly influenced by the ordinary language use of the word to signify "long-lasting" or "having a substantial impact". Third, it can be a subtopic of the broader sustainability discourse. The latter can encompass the second option, whereas the inverse relation does not hold.

One way to look at the relation between "fiscal" and "sustainability" is by asking in what respect "public debt" is a sustainability topic. In answering this question, we move towards questions of intergenerational justice. To show this, consider the fact that there are many severe crises which have little to do with sustainability since they do not involve intergenerational equity. For example, a volcanic eruption or a famine resulting from catastrophic weather conditions indeed causes much harm, and emergency plans are of utmost importance in responding to and managing such crises. Humanitarian aid has to be organised. However, there is little to nothing that the sustainability discourse can contribute to informing the necessary decision-making in such cases. To bring sustainability, as so far discussed, into crisis management would require *at least* a future-oriented perspective that includes intergenerational issues, involving potential harm for future generations or issues of an unjust distribution of benefits and burdens between generations. Questions and issues surrounding intergenerational justice in general are not totally new societal topics. Pension systems or responsibility for one's children, for example, involves intergenerational considerations. However, from the perspective of Pearce and Atkinson (1998), sustainability goes beyond "the living together of the currently living generations" (=intergenerational justice$_1$); it involves at least some of the not-yet-living generations (intergenerational justice$_2$). Public debt does raise concerns about intergenerational justice in general as the current living generation profits on the shoulders of its own children, who will have to pay back the debt, that is, they do not profit but will carry the burden. However, this is intergenerational justice$_1$, not intergenerational justice$_2$.

There are at least two arguments for looking at public debt from the perspective of intergenerational justice$_2$. First, the three constitutive features introduced above do not characterise "sustainability" in terms of "environmental risks" but more generally in terms of "system risks". Many environmental risks, today in particular ones related to emissions and the growing fragility of ecosystems, belong to the set of system risks. "System risk" refers to the probability of an event occurring that provokes a major system breakdown which goes beyond the given resilience threshold. A nuclear war, for example, is a severe system risk. Struck by such an event, the system in question will no longer be able to deliver the relevant functions. This holds for both human and natural systems. Hence, public debt can belong to the realm of intergenerational justice$_2$ if it has a relevant probability of triggering an event that provokes a system breakdown, where "system" here refers to human societies.

Second, intergenerational justice$_2$ comes into an evaluation scheme if there are relevant concerns about the opportunity space for future generations to realise a good life. Imagine the following. Unfortunately, climate change progresses faster than expected and a nation faces unexpected early severe challenges to deal with flooding. Fortunately, there is still enough time to adapt to the changing environment, at least if investments in flood management can be realised. Unfortunately, however, and due to a critical state of the public finances, these investments could only be made by either surpassing a critical threshold of public debt or by reducing social welfare contributions (or some other important expenditure). The first option could provoke a financial system breakdown; the second a social system breakdown.

For these two reasons, public debt should indeed also be evaluated from a broader sustainability perspective. Public debt is related to opening or closing opportunity spaces for both those currently living and future generations to realise a decent life. In this respect, it involves issues of intergenerational justice$_2$, because it is not only related to burdens and benefits for generation currently alive, but to potential opportunities for realising wellbeing for future generations not yet alive. More directly, the topic is related to (a) the possibility of a major (societal) system breakdown transcending the resilience level and (b) the capacity of societies to deal with a changing environment or new challenges.

In what respect these risks exist is an empirical question. Answering it would go much beyond the scope and expertise of the present chapter. However, the line of argument presented leads to the conclusion that there are two substantial instrumental goods involved in a serious discussion of fiscal sustainability. The first is resilience of the social systems in question. A financial breakdown can lead to a system breakdown. Most importantly,

today's debts can have that effect in an imagined future. The second is society's potential to act and react for transformation and adaptation. Financial shortages due to severe public debt can substantially impact on innovation and adaptation processes. Both a system breakdown and a restricted opportunity space for innovation and adaptation will most certainly have negative impacts on the wellbeing of the generation affected. Hence, by including the instrumental goods "resilience of the social system" and "societal action and reaction potential" there is a well-founded bridge linking the traditional sectorial debate on fiscal risks with the much broader sustainability discourse.

As for measurement, methodologically speaking there is nothing special to it. What has been said in section "Sustainability Assessment" also holds true for "fiscal sustainability" with any of the possible meanings of "sustainable". First, assessing the risks related to public debt is always dependent on what Renn and Klinke (2015) call "socio-political ambiguity" and what I have referred to as theoretical elements characterising the "if" in the related "if-then" statements. These elements include all those in sections "Sustainable Development and the Range of Conceptions of Scientific Sustainability" and "Sustainability Assessment". Searching for an aggregate number, like the ratio between debt and GDP, can be informative and useful, but it only expresses one perspective with a specific reduction of complexity. In line with Hopton et al. (2010), we should claim that complementary approaches are needed and that decision-making should not rely on only one specific approach. Second, assessing the risks related to public debt within a sustainability framework should refrain from looking at it from a sectorial perspective. The fiscal situation of a state has to be related to other relevant societal challenges (e.g. equity, democratic legitimacy, environmental challenges, and so on) and become part of considerations on the societal potential to act and react to cope with its changing environments.

Conclusion

It is reasonable to look on fiscal sustainability as a specific subtopic of sustainable development. It deals with states' financial capital as a specific type of resource needed to ensure the wellbeing of the currently living and future generations. However, we can hardly expect to find something like a well-defined tipping point or a threshold level as a criterion for sustainability on which we can base decisions unequivocally. Fiscal sustainability is part of a complex equation also involving natural resources, system fragility, human

wellbeing, efficiency and the requirements of democracy and business, and so on. Any solution will further depend on additional conceptual investments, as argued in this chapter. Accordingly, we should aim to have a framework for finding reasonably balanced solutions rather than strive for optimal solutions.

Notes

1. It is not within the scope of this chapter to distinguish between sustainable development and sustainability.
2. In an enquiry on existing research on SD at universities and other research institutions in the tri-national Upper Rhine Region (Burger et al. 2014), we distinguished 18 different topic fields ranging from energy to water issues to household consumption to governance and transition. We also included the field of fiscal sustainability.
3. The famous *woodcutter rule* of not harvesting more wood from a forest than will grow again was first written down by Hans Carl von Carlowitz in 1713 (von Carlowitz 2012). Its roots go back to the mediaeval period as Ulrich Grober (2010) reveals. However, it is standard in today's scientific reasoning on SD that the woodcutter rule is much too simple to serve as a foundation for conceptualising SD.
4. The WCED's definition of sustainable development uses the term "needs" (satisfaction of needs) to address quality of life. For reasons beyond the scope of this chapter, "needs" will not be considered; instead, the chapter will refer in general to a good or decent life, leaving aside the question of an appropriate metric for wellbeing.
5. Among other things, this includes environmental risks.
6. Neither scarcity of resources nor intergenerational issues are really new topics, cf. for example Solow (1974). However, I follow Pearce and Atkinson (1998) and many others in saying that SD along the lines of the WCED report brings in new aspects.
7. The famous controversy about weak and strong sustainability, that is, about substitution of especially natural resources (capital) by economic or social capital often builds on such a capital stock approach.
8. One could add the famous three-pillar approach to this list. However, I do not look upon it as a theoretical approach to SD but rather as a pragmatic management rule: consider at least societal, environmental, and economic issues when dealing with the consequences of our decisions.
9. Cf. the nice metaphor in Pearce et al. (1991, p. 3): SD could be looked upon as something you simply have to like just like motherhood or apple pie.
10. Meyer deals with intergenerational justice in general, not specifically SD.

11. There are further distinguishing elements such as how to conceptualise future generations (e.g. Partridge 2001) or how to cope with the related non-identity problem (Parfit 1984). Discussing these would, however, go beyond the scope of this chapter.
12. Instead of using the unclearly defined notion of a "failed state", Call 2011 discusses three fundamental functions a state has to fulfil: providing its population with basic goods; infrastructure security; and a legitimacy of political elites.
13. Additionally, there is a vast body of literature on ecosystem services and on environmental management.
14. To give an example: if you include an indicator such as "number of deaths per KWh produced", nuclear power will be rated top since there are few (direct) casualties (e.g. Chernobyl). If you take an indicator such as "number of persons evacuated per KWh produced", the rating will be inverted, given the numbers from Chernobyl and Fukushima.
15. Further elaborating this argument would be the topic of another chapter.

References

Adger, W. N. (2006). Vulnerability. *Global Environmental Change, 16*(3), 268–281.

Armitage, D. R., Plummer, R., Berkes, F., Arthur, R. I., Charles, A. T., Davidson-Hunt, I. J … Wollenberg, E. K. (2009). Adaptive Co-management for Social-Ecological Complexity. *Frontiers in Ecology and the Environment*, *7*(2), 95–102.

Barry, B. (1999). Sustainability and Intergenerational Justice. In A. Dobson (Ed.), *Fairness and Futurity. Essays on Environmental Sustainability and Social Justice* (pp. 93–117). Oxford: Oxford University Press.

Bellagio Principles. (1997). *Bellagio Principles for Sustainability Assessment*. Winnipeg: International Institute for Sustainable Development.

Berkes, F., Colding, J., & Folke, C. (Eds.). (2006). *Navigating Social-Ecological Systems. Building Resilience for Complexity and Change*. Cambridge: Cambridge University Press.

Burger, P., & Christen, M. (2011). Towards a Capability Approach of Sustainability. *Journal of Cleaner Production, 19*(8), 787–795.

Burger, P., Hofstetter, N., Ott, J., & Riesen, J. (2014). *Nachhaltigkeitsforschung in der Trinationalen Metropolregion Oberrhein (TMO) / Recherche sur la durabilité dans la Région Métropolitaine Trinationale du Rhin Supérieur (RMT)*. Basel: University of Basel.

Call, C. T. (2011). Beyond the "Failed State": Toward Conceptual Alternatives. *European Journal of International Relations, 17*(2), 303–326.

Chambers, R., & Conway, G. R. (1992). *Sustainable Rural Livelihoods: Practical Concepts for the 21st Century* (Institute of Development Studies, Discussion Paper 296). Brighton: Institute of Development Studies. University of Sussex.

Christen, M. (2013). *Die Idee der Nachhaltigkeit. Eine werttheoretische Fundierung*. Marburg: Metropolis.

Christen, M., & Schmidt, S. (2011). A Formal Framework for Conceptions of Sustainability – a Theoretical Contribution to the Discourse in Sustainable Development. *Sustainable Development, 20*(6), 400–410. doi:10.1002/sd.518.

Dobson, A. (1996). Environmental Sustainabilities: An Analysis and a Typology. *Environmental Politics, 5*(3), 401–428.

Ehrenfeld, J. R. (2004). Can Industrial Ecology Be the "Science of Sustainability"? *Journal of Industrial Ecology, 8*(1–2), 1–3.

Farrelly, C. (2007). Justice in Ideal Theory: A Refutation. *Political Studies, 55*(4), 844–864.

Fischer-Kowalski, M., & Haberl, H. (2007). *Socioecological Transitions and Global Change. Trajectories of Social Metabolism and Land Use*. Cheltenham: Edward Elgar.

Gallopin, G. C. (2006). Linkages between Vulnerability, Resilience, and Adaptive Capacity. *Global Environmental Change, 16*(3), 293–303.

Global Footprint Network. (2010). *Data and Results from the 2010 Edition*. Oakland: Global Footprint Network. Retrieved from http://www.footprintnetwork.org/en/index.php/GFN/page/footprint_data_and_results

Government of Kanton Basel-Stadt. (2014, September 18). *Press Release on Budget 2015* [Press Release]. Retrieved from http://www.medien.bs.ch/news/2014-09-18-mm-61062.html).

Gowdy, J. (2005). Toward a New Welfare Economics for Sustainability. *Ecological Economics, 53*(2), 211–222.

Grober, U. (2010). *Die Entdeckung der Nachhaltigkeit. Kulturgeschichte eines Begriffs*. München: Antje Kunstmann.

Grunwald, A. (2009). Konzepte nachhaltiger Entwicklung vergleichen – aber wie? Diskursebenen und Vergleichsmaßstäbe. In T. Egan-Krieger, J. Schulz, P. P. Thapa, & L. Voget (Eds.), *Die Greifswalder Theorie starker Nachhaltigkeit. Ausbau, Anwendung und Kritik* (pp. 41–64). Marburg: Metropolis.

Habermas, J. (2007). *Theory of Communicative Action*. Cambridge: Polity Press.

Hopton, M. E., Cabezas, H., Campbell, D., Eason, T., Garmestani, A. S., Heberling, M. T., et al. (2010). Development of a Multidisciplinary Approach to Assess Regional Sustainability. *The International Journal of Sustainable Development & World Ecology, 17*(1), 48–56.

Janssen, M., & Ostrom, E. (2006). Resilience, Vulnerability, and Adaptation: A Cross-Cutting Theme of the International Human Dimensions Programme on Global Environmental Change [Editorial]. *Global Environmental Change, 16*(3), 237–239.

Kates, R. W., Parris, T. M., & Leiserowitz, A. A. (2005). What Is Sustainable Development? Goals, Indicators, Values, and Practice. *Environment: Science and Policy for Sustainable Development, 47*(3), 8–21.

Kemp, R., Loorbach, D., & Rotmans, J. (2007a). Transition Management as a Model for Managing Processes of Co-evolution. *The International Journal of Sustainable Development and World Ecology* (special issue on (co)-evolutionary approach to sustainable development)*, 14*, 78–91.

Lafferty, W. M. (2004a). Introduction: Form and Function in Governance for Sustainable Development. In W. M. Lafferty (Ed.), *Governance for Sustainable Development: The Challenge of Adapting Form to Function* (pp. 1–31). Cheltenham & Northampton: Edward Elgar.

Lafferty, W. M. (2004b). From Environmental Protection to Sustainable Development: the Challenge of Decoupling Through Sectoral Integration. In W. M. Lafferty (Ed.), *Governance for Sustainable Development: The Challenge of Adapting Form to Function* (pp. 191–220). Cheltenham & Northampton: Edward Elgar.

Meadowcroft, J. (2004). Participation and Sustainable Development: Modes of Citizen, Community and Organisational Involvement. In W. M. Lafferty (Ed.), *Governance for Sustainable Development: The Challenge of Adapting Form to Function* (pp. 161–190). Cheltenham & Northampton: Edward Elgar.

Meyer, L. (2015). Intergenerational Justice. In *Stanford Encyclopedia of Philosophy*. Retrieved March 12, 2017, from https://plato.stanford.edu/entries/justice-intergenerational/

Moser, A. (2014). *Ein Metaverständnis der Nachhaltigkeitsbewertung. Für eine vergleichende strukturelle Analyse zur Standortbestimmung der wissenschaftlichen Nachhaltigkeitsbewertung im Energiebereich* (Master Thesis). Basel: University of Basel.

Norton, B. G. (2005). *Sustainability. A Philosophy of Adaptive Ecosystem Management*. Chicago: The University of Chicago Press.

Nuclear Energy Agency. (2000). *Nuclear Energy in a Sustainable Development Perspective*. Paris: OECD. Retrieved from https://www.oecd-nea.org/ndd/docs/2000/nddsustdev.pdf.

Ostrom, E. (2009). A General Framework for Analyzing Sustainability of Social-Ecological Systems. *Science, 325*, 419–422.

Ott, K., & Döring, R. (2008). *Theorie und Praxis starker Nachhaltigkeit*. Marburg: Metropolis.

Pahl-Wostl, C. (2009). A Conceptual Framework for Analysing Adaptive Capacity and Multi-Level Learning Processes in Resource Governance Regimes. *Global Environmental Change, 19*(3), 354–365.

Parfit, D. (1984). *Reasons and Persons*. Oxford: Oxford University Press.

Partridge, E. (2001). Future Generations. In D. Jamieson (Ed.), *A Companion to Environmental Philosophy* (pp. 377–389). Oxford: Blackwell Publisher.

Pearce, D., & Atkinson, G. (1998). *The Concept of Sustainable Development: An Evaluation of Its Usefulness Ten Years after Brundtland* (Center for Social and Economic Research on the Global Environment, Working Paper PA 98-02). Norwich: University of East Anglia.

Pearce, D., Markandya, A., & Barbier, E. B. (1991). *Blueprint for a Green Economy*. London: Earthscan Publications.

Rauschmayer, F., Omann, I., & Frühmann, J. (Eds.). (2012). *Sustainable Development: Capabilities, Needs, and Well-Being*. Abingdon: Routledge.

Renn, O., & Klinke, A. (2015). Risk Governance and Resilience: New Approaches to Cope with Uncertainty and Ambiguity. In U. F. Paleo (Ed.), *Risk Governance* (pp. 19–41). Dordrecht: Springer.

Robèrt, K. H., Schmidt-Bleek, B., De Larderel, J. A., Basile, G., Jansen, J. L., Kuehr, R., et al. (2002). Strategic Sustainable Development – Selection, Design and Synergies of Applied Tools. *Journal of Cleaner Production, 10*(3), 197–214.

Smith, A., & Sterling A. (2008*). Social-Ecological Resilience and Socio-Technological Transitions: Critical Issues for Sustainability Governance* (Social, Technological and Environmental Pathways to Sustainability Working Paper 8). Brighton: Social, Technological and Environmental Pathways to Sustainability Centre.

Solow, R. M. (1974). Intergenerational Equity and Exhaustible Resources. *The Review of Economic Studies, 41*, 29–45.

Sustainable Development Commission UK. (2006). *The Role of Nuclear Power in a Low Carbon Economy* (UK Sustainable Development Commission Position Paper). London: UK Sustainable Development Commission. Retrieved from http://www.sdcommission.org.uk/publications.php?id=344

United Nations. (1992). *The Rio Declaration*. Rio de Janeiro: United Nations Conference on Environment and Development. Retrieved from http://www.un.org/documents/ga/conf151/aconf15126-1annex1.htm.

von Carlowitz, H. C. (2012). *Sylvicultura oeconomica. Anweisung zur wilden Baum-Zucht (Faksimile der Erstauflage)*. Remagen: Kessel Verlag.

Voss, J. P., Bauknecht, D., & Kemp, R. (Eds.). (2006). *Reflexive Governance for Sustainable Development*. Cheltenham & Northampton: Edward Elgar Publishing.

Voss, J. P., Newig, J., Kastens, B., Monstadt, J., & Nölting, B. (2007). Steering for Sustainable Development: A Typology of Problems and Strategies with Respect to Ambivalence, Uncertainty and Distributed Power. *Journal of Environmental Policy & Planning, 9*(3–4), 193–212.

Weale, A. (2011). New Modes of Governance, Political Accountability and Public Reason. *Government and Opposition, 46*(1), 58–80.

World Commission on Environment and Development. (1987). *Our Common Future*. New York/Oxford: Oxford University Press.

7

Measuring Governance: Revisiting the Uses of Corruption and Transparency Indicators

Robert P. Beschel Jr

Introduction

Humanity has long grappled with the challenge of measuring corruption and understanding its trends and impact. Italian history provides a particularly interesting case study. In his incisive and provocative study, "Corruption and the Decline of Rome", historian Ramsay MacMullen argues that by the late empire a massive shift had occurred in government ethics and civic life. Corruption had become a cancer eating away at the sinews of the state:

> Bribery and abuses always occurred, of course. But by the fourth and fifth centuries they had become the norm: no longer abuses of a system, but an alternative system in itself. The cash nexus overrode all other ties. Everything was bought and sold: public offices including army commands and bishoprics, judges' verdicts, tax assessments, access to authority on every level, and particularly the Emperor. The traditional web of obligations became a marketplace of power, ruled only by naked self-interest. Government's operation was permanently, massively distorted. Imperial authority was of course upheld, since it was

This chapter was prepared by Robert P. Beschel Jr. in the context of a workshop sponsored by the European University Institute and the Robert Schuman Centre for Advanced Studies, *Workshop on Global Governance by Indicators: Measuring Corruption and Corruption Indices*, which took place in Florence in October 2013. It has been revised and updated to reflect subsequent developments. The views within it are those of the author and do not represent the World Bank and its affiliates.

R.P. Beschel Jr (✉)
World Bank, Washington, DC, USA

D.V. Malito et al. (eds.), *The Palgrave Handbook of Indicators in Global Governance*,
https://doi.org/10.1007/978-3-319-62707-6_7

> precisely the source of illicit gain. But its power was dissipated into thousands of private channels in a way that did not happen in the earlier empire. (MacMullen 1988; Williams n.d.)

While the loss of Rome's geopolitical influence was real and had been under way for well over a century prior to the final deposing of Emperor Romulus in 476 AD, the role of corruption in bringing it about remains more contentious. It is clear that the empire that was able to protect itself and project power effectively under Trajan—and to mobilise the political, military, financial and administrative resources that would allow it to—had become incapable of doing so by the late fifth century. Was the decline in virtue and the growing abuse of public office real? Or is this a post-hoc rationalisation and only perceived as such against some idealised vision of the past? While MacMullen makes a powerful and compelling case, his argument falls more into the realm of historical interpretation than established fact. Sixteen centuries later, it is virtually impossible to conclusively prove that corruption in the late Roman Empire was qualitatively and quantitatively different to that in earlier periods. We lack access to the sort of data and benchmarks that would allow definitive judgements to be made.

More recently, contemporary scholars have been working hard to take advantage of the techniques available to modern social science to address issues of corruption with more accuracy and rigor. Since the late 1980s, efforts to apply a variety of analytical techniques to the empirical study of corruption have gathered particular momentum. As Box 7.1 indicates, a wealth of comparative indices and assessment tools are now available to everybody with an internet connection, and additional proprietary assessments and analyses are available to those who have access to or are otherwise willing to pay for them. Some indices, such as Transparency International's (TI) Corruption Perceptions Index (CPI), seek to directly measure corruption and provide international comparative data. Others, such as the Bertelsmann Stiftung's Transformation Index (BTI), focus more on broader governance issues and only address corruption implicitly. Still other indicators, such as the Public Expenditure and Financial Accountability (PEFA) and the Business Environment and Enterprise Performance Survey (BEEPS) indicators, seek to measure the quality of selected government functions that will have a direct bearing on certain dimensions of corruption. (PEFA was not originally intended to provide cross-country comparative data, but is increasingly being used in this context.) Our discussion will be directed towards these indicators, particularly those that are publicly available.

Box 7.1 Commonly Used Global Indices and Metrics for Corruption

Perception-Based Indices: Transparency International Global Corruption Barometer; World Values Survey; Gallup Surveys; Business Environment and Enterprise Performance Survey (BEEPS); Afrobarometer surveys.

Expert-Based Indices and Assessments: Transparency International Corruption Perceptions Index (CPI); Global Integrity Reports; World Bank Country Policy and Institutional Assessment (CPIA); Public Expenditure and Financial Accountability (PEFA) Assessments; World Economic Forum Global Competitiveness Report; United Nations Convention Against Corruption Assessments; Global Financial Integrity Illicit Financial Flows assessments.

Composite Indices: Worldwide Governance Indicators (WGIs), Bertelsmann Stiftung Transformation Index (BTI); World Justice Project Rule of Law Index.

Source: Own compilation

This Box only focuses on the most widely known and publicly available indices. A variety of other indices are produced by private firms—such as the Economist Intelligence Unit, Verisk Maplecroft, RSM and the PRS Group—and are often tailored to the needs of certain clients or sectors

The above efforts have found receptive audiences. Businesses have used them as inputs in investment decisions. Governments have alternately used them as a source of pride and/or motivation or criticised them as biased or inaccurate. Civil society organisations have used them as tools to advance their political and social causes. Donor agencies, such as the World Bank, have sought to use them both to guide resource allocation and to inform country programming and support.

Are these indicators ultimately capable of carrying this freight? This chapter will review their uses and evolution, with particular reference to those used by the World Bank. It will reflect on the goals of these indicators and the types of use that they are typically put to. It will highlight the relative strengths and weaknesses of the indicators most in circulation. It will provide some context on how they have evolved and what directions they are likely to head in in the next decade. In addition, it will both begin and conclude with a few reflections on how they should best be used.

Uses of Corruption and Transparency Indices

Before probing the efficacy of these indicators, it is important to inquire into their intended uses. Addressing this question is complicated, as different actors typically have different objectives, expectations, time and capacity to absorb and understand the indices. There is seldom any "one size fits all"

approach that will satisfy the disparate demands of different constituencies, but the uses of corruption and transparency indices typically fall into six categories:

1. *To evaluate the nature and depth of a problem by placing it in a global context.* At the most basic level, corruption-related indicators reflect the desire of many different actors to simply know how "good" (or more typically, how bad) the governance environment is in a given state or polity. This desire in turn feeds a hunger for a clear set of rankings that can summarise complex phenomena into a simple number or readily comprehendible benchmark that can be instantly grasped and understood by a broad proportion of the population.
2. *To provide guidance as to what potential solutions should look like*. Ideally, indicators should be both "actionable" (in that they point the way for those interested in taking concrete steps to make things better) and "action-worthy" (in that they focus attention on important measures that, if enacted, will have an impact). "Black box" indicators that are non-transparent or rely on proprietary methodologies and assessments are problematic in this respect.
3. *To monitor progress and performance.* In addition to being actionable, a robust set of indicators should ideally allow both internal and external observers to monitor progress over time and assess whether a given government is making progress in the struggle against corruption. Towards this end, any assessments—and the rankings that flow from them—should be transparent and replicable by different observers. They should draw on stable and predictable data sources and avoid any bias in perception or analysis.
4. *To shed light on underlying dynamics and drivers*. Governance indices are often utilised as inputs in a range of academic analyses that seek to probe the efficacy of various factors and causal relationships. To best serve the scholarly and research communities, both the indices and their underlying methodologies should be widely accessible. They should also facilitate ease of data collection by relying as much as possible on existing data and approaches that are readily accessible.
5. *To inform decision-making by external parties.* In addition to those interested in their empirical and theoretical value, various constituencies may rely on indices in whole or in part to inform important decisions. Businesses can use them as an input in investment decisions or risk-mitigation measures. Donors may use them when allocating resources or when making country programming decisions.

6. *To create incentives for positive change.* Governance indices are rarely strictly empirical and/or technical in nature. There is often a strong normative dimension that either implicitly or explicitly seeks to use the indicators to create impetus for positive change. The goal is not just to accurately reflect reality, but to create incentives that will ultimately help to reduce corruption, enhance transparency and improve the quality of institutions.

In addition to the attributes highlighted above, an ideal set of governance indicators should embody other virtues. They should be politically understandable and acceptable to a wide range of governments and their citizens, so that as many countries as possible are willing to embrace and push for them and not reject them as ideologically undesirable or culturally inappropriate. Their methodologies should be readily replicable so that different groups using them arrive at similar conclusions. Moreover, they should avoid what some scholars have characterised as "isomorphic mimicry"—a tendency to adopt the form but not the functionality when transferring legal and institutional reforms from one context to another (Pritchett et al. 2010).

Obviously, it is nearly impossible for an indicator or set of indicators to meet all of these differing objectives and expectations. As will be discussed in greater detail below, indicators that are simple and easily accessible—such as Transparency International's CPI or the Brookings Institution's World Governance Indicators (WGIs)—typically lack enough depth and nuance to provide much direct guidance to policymakers. Indicators that rely on expert judgements and assessments, such as those reviewing the efficacy of anticorruption legislation or institutions, may produce complex and nuanced judgements that are not easily summarised or disseminated. Indices that donor agencies rely upon in resource allocation decisions, such as the World Bank's Country Policy and Institutional Assessment (CPIA) indicators, may not cover the issues that are most germane for textile or extractive firms, for example, weighing up whether to invest in a given country.

Conceptual, Methodological and Practical Challenges Associated with Comparative Indices

In addition to the difficulty of trying to satisfy the diverse needs of different constituencies, many global indices confront a host of conceptual, methodological and practical challenges that certainly complicate—and some would claim compromise—their utility (see Thomas 2010). Several of these are described below.

Conceptual Challenges Conceptually, our intuitive understanding of key governance concepts such as "corruption" or the "rule of law" often breaks down when efforts are made to operationalise them. Corruption is a shorthand term that can refer to dozens if not hundreds of administrative pathologies. The *Arthashastra*, believed to have been compiled in India between 200 BC and 300 AD, for example, lists over 40 types of corrupt and illicit behaviour. (The sanctions could range from death through torture for the most severe crimes to public humiliation and smearing with cow dung and ashes for misdemeanours). Today, "corruption" can cover anything from collusion in procurement to nepotism in recruitment, to paying speed money and making petty facilitation payments, to extortion and protection payments, to the granting of licenses and favourable access for political contributions—and a good deal more besides. Some variants are extractive; others are collusive. Some may be widespread, others relatively rare. Their cost and impact on the body politic can vary greatly.

Of particular concern is the distinction between petty and grand corruption, with the latter being much less visible and significantly more difficult to prove. Many perception-based indicators are likely to capture public views about petty corruption, which is more germane to an individual's daily experience, than grand corruption, although the latter could in many cases end up having a significantly more pernicious influence. Tunisia, for example, ranked a respectable 59 in TI's CPI index, in 2010, placing it in the upper tier of Middle East and North Africa (MENA) countries in the struggle against corruption immediately prior to Ben Ali's ousting at the onset of the Arab Spring. Yet more detailed analysis after the revolution revealed that Ben Ali's family controlled around 220 firms responsible for around 32 per cent of all private sector profits, even though they accounted for only 3 per cent of private sector output (Rijkers et al. 2014). More perniciously, the analysis revealed a close nexus of regulatory capture among the connected firms, particularly with regard to allowing entry by other players into tightly regulated markets.

Similar definitional and conceptual challenges confound efforts to analyse other principles of good governance. Does "rule of law" refer to the quality of the underlying legislation, the ability of citizens to gain access to the courts and receive judicial redress within a reasonable period of time, or the consistency and predictability with which laws are enforced? Does it apply to the enforcement of contracts between private parties, the willingness of the police and prosecutors to abide by due process restraints, or the security of individuals and property from violent crime? If, as one would expect, all of these dimensions are important to some measure, how should they best be combined and weighted?

The reality is that the discipline of political science currently offers little true guidance on these issues. There is no overarching "theory of corruption" or "theory of good governance" to provide heuristic guidance, nor is there a robust set of empirical evidence pointing clearly in a given direction. We can point to relatively robust correlations in macro comparative data between benefits that we value (such as high or rapidly growing levels of GDP) and the principles we believe align with "good governance" (such as low public perceptions of corruption), but the causality is often difficult to tease out—particularly when we delve deeper and seek to unpack the impacts of various sub-components. We can cull lessons from individual country experiences about the cost of different types of corruption and which interventions worked well or did not (and we are increasingly doing so), but this raises questions of impact, scalability and replicability elsewhere. At this stage, we often do not know which sub-components are most important (or pernicious) and the proportions in which they are present and interact within a given polity. Nor do we know whether these judgements will hold true for virtually all developing countries or for significant sub-sets of countries at particular stages of development, or within certain administrative traditions, or that they may have worked for a lucky few, but it is all ultimately idiosyncratic and heavily context-specific.

Methodological Challenges Beyond these broader conceptual questions, the first generation of governance indicators and assessments—including the CPI and WGI indicators—typically suffered from a number of methodological challenges that further complicated their use (see Arndt and Oman 2006). The first was a high degree of correlation among the underlying source data. One would expect, for example, questions involving corruption and the rule of law to be highly inversely correlated. Countries that scored high on indices of corruption would do poorly on those measuring rule of law, and vice versa. One would also expect various sub-dimensions of broader factors such as rule of law to be closely correlated, for instance, public perceptions regarding the functioning of the police force and the security of property rights. One would anticipate a fair degree of correlation between perception-based indices and those developed by experts, as the latter can be influenced by polling data on the prevalence of corruption, and the former can be influenced by expert assessments that corruption within a given country is a problem—a topic that we will return to later in the paper.

A second challenge is the extent to which data are comparable over time. Perhaps the greatest use, and most frequent abuse, comes with TI's Corruption

Perceptions Index, where it is common for the press to interpret movement up or down the ranking as a sign that a country is making progress (or failing to do so) in the broader struggle against corruption. In December 2014, for example, Pakistan's "Dawn" newspaper—the oldest and most widely read English language paper in the country—noted that Pakistan's ranking in the CPI corruption index had improved from 127 out of 177 to 126 out of 175 (Ahmed 2014). In reality, such changes are likely to be meaningless and due more to statistical noise than any significant change in perceptions, let alone a change in actual levels of corruption. Prior to 2012, TI's annual CPI rankings were sui generis and could not be benchmarked against scores from previous years (see CPI 2012).[1]

Others have accused these rankings with bias in their samples of sources, noting that they tend to have a private sector bias or to focus disproportionately on business surveys or expert judgements (Donachev and Ujhelyi 2014). Such indices can be sensitive to absolute levels of corruption (such as the number of occurrences) rather than relative corruption levels (percentage of the population affected), which implies that perceptions will tend to be biased upwards for larger countries. Indices may exhibit diminished sensitivity to corruption experience, which would mean that they may do better in low-corruption countries than in high-corruption contexts. Some argue that the experience of ordinary citizens or the poor and the disenfranchised is often undervalued. Such surveys can also end up having a closed-loop, mutually reinforcing dynamic. As one expert has lamented, "we are basically talking to ourselves" (Arndt and Oman 2006, p. 50).

Another challenge is that of confounding or unmeasured variables. Returning to the rule of law illustration above, it could be that both corruption and the rule of law are strongly influenced by a third variable—the quality of political leadership—which is not being explicitly tracked and monitored. Fortunately, social science offers a variety of research design approaches that can help to reduce or eliminate such problems, although they are not without their trade-offs and drawbacks.

The result of these methodological challenges is to expand the error margins for various indices to the point that makes it difficult to attribute much significance to modest changes in country rankings from year to year. Some assessments, such as the WGI, have sought to address this challenge by providing confidence intervals where users could be sure that there was a 90 per cent probability or greater that a given country would fall within a given bandwidth, with the breadth of these intervals varying both between dimensions and countries. However, critics allege that even these bandwidths do not fully capture the potential variation entailing from the methodological challenges cited above (Arndt and Oman 2006).

Practical Challenges Beyond the conceptual and methodological challenges highlighted above, efforts to develop robust governance indicators and assessment tools often suffer from a variety of practical problems. Access to accurate, timely and reliable data is a chronic challenge, particularly in smaller and less developed countries or in autocratic societies where potentially unflattering assessments are not welcome. For many NGOs active in producing governance assessments, the availability of adequate funding is another chronic problem.

One of the most serious criticisms levelled against cross-country comparative indicators is that they are simply not actionable. A well-intentioned new Minister of Justice who is concerned that his or her country's ranking in the WGI's rule of law indicators is not where it should be receives virtually no guidance as to what can be done to improve its standing. The WGI is a composite index that draws on four different types of source data: surveys of households and firms, commercial business information providers, non-governmental organisations and public sector organisations (including the World Bank). Various inputs are assigned to each of the WGI's six indicators, they are rescaled to run from 0 to 1, and they are adjusted using an unobserved components model to construct a weighted average of the individual indicators. The result is a number that can be benchmarked against regional and global averages. However, it cannot be readily deconstructed into its component parts so that reformers can know where they are falling short and why to take remedial measures.

Historical Evolution

Over the past two decades, in response to both popular demand and the need to address some of the challenges identified above, global governance indices and assessments have moved in two general directions. The first has been a push to develop more accurate, robust and comprehensive comparative indices. The second has been to develop more narrow technical assessments of specific government institutions and functions. Both will be discussed below.

The first major global corruption index and the one that remains by far the most well known today is Transparency International's Corruption Perception Index, or CPI. Begun in 1995, it currently covers around 170 countries and assesses "the degree to which corruption is perceived to exist among public officials and politicians" (Transparency International 2012). The index relies on a mix of expert interviews and surveys. Critics have argued that it suffers from a number of the flaws highlighted above, yet the general consensus appears to be that it is useful in focusing broader political attention on the

problem of corruption (Hough 2016). TI sought to augment the CPI with the "Global Corruption Barometer", which is also a perception-based index. However, it differs from the CPI, in that it is based on a direct survey of individuals. The most recent version, conducted in 2013, covered 95 countries and asked direct questions such as "have you ever been asked to pay a bribe?" Sampling errors, typically in the range of 3 per cent, are relatively modest.

Another long-standing governance index, the Worldwide Governance Indicators, has many features in common with TI's CPI Index. (It was started in 2002 based on ongoing research at the World Bank Institute, and it has subsequently moved over to the Brookings Institution.) The WGIs are also a composite index that draw upon a large number of public and private sources to produce comparative global rankings. The indicators cover a broader range of governance topics than the CPI, including voice and accountability, political stability, lack of violence, government effectiveness, regulatory quality, rule of law and control of corruption. It also covers a broader range of countries (over 200).

The WGIs have also been subject to many of the standard criticisms, not the least of which is that of construct validity—the extent to which the measures used actually reflect the concepts that they are purporting to measure (GSDRC 2010). Daniel Kauffmann and Aart Kraay, two of the principal creators of the WGIs, have responded forcefully to these criticisms (Kaufmann and Kraay 2007). While acknowledging the validity of many of the arguments arrayed against them, they also maintain that the exercise of developing such indices remains both worthwhile and inherently problematic. They argue that all governance indicators have inherent weaknesses, that there are no easy solutions in measuring governance and that the links between governance and broader development outcomes are complex. In their view, global indices, such as the WGIs, are complementary to more detailed and actionable approaches.

Newer indices, such as the World Justice Project's Rule of Law Index, have sought to utilise composite methodologies that combine surveys and expert assessments. Created in 2012, the Rule of Law Index seeks to assess a variety of topics ranging from constraints on government powers to order and security, and to civil and criminal justice (World Justice Project 2017). It also addresses absence of corruption. As a result of variations in the number of countries analysed and adjustments in data sources and methodology, its results are not strictly comparable over time—a common problem with such indices. Both the CPI and the WGIs have also been adapted over the years to enhance the comparability and robustness of their results. As noted above, the greatest change to the CPI came in 2012, when efforts were made to

standardise the country rankings on a scale of 0–100 and to include single-year data from each data source, which was intended to facilitate comparisons over time (see Transparency International Ukraine 2012).

At the same time that global indices were evolving, other efforts were being made to develop more granular comparative assessments of various dimensions of government performance. These "second generation" assessments were often expert-based analyses that followed standard formats and methodologies and were significantly more detailed than broader indices. Many were developed for purposes beyond measuring and monitoring corruption, such as giving donors confidence that they could support national budgetary systems in providing aid, or reviewing compliance with international treaty obligations. However, they also play an important role in developing qualitative assessments of the legal and institutional frameworks that countries use to promote integrity and combat illicit behaviour. Three are particularly important for our purposes: (1) the Public Expenditure and Financial Accountability (PEFA) assessments; (2) the United Nations Convention Against Corruption (UNCAC) Assessments; and (3) the monitoring of illicit financial flows.

Of the various assessments referenced above, the PEFA assessment is probably the broadest and most well established. As of late 2016, nearly 550 assessments had been conducted worldwide. PEFA has recently updated its methodology and focuses on improving the quality of budget and financial management. It reviews seven criteria along the standard budget cycle: budget reliability, the transparency of finances, the management of assets and liabilities, policy-based fiscal strategy and budgeting, predictability and control in budget execution, accounting and reporting and external scrutiny and audit. Beneath these seven pillars, the PEFA framework uses 31 specific indicators disaggregated into 94 dimensions that focus on key measurable aspects of the public financial management (PFM) system, such as actual versus predicted budget outcomes.

The PEFA assessments are supported by a number of different donors and are widely viewed as the standard against which PFM systems are measured. The assessments are broken down into an "A" to "D" ranking system that is often used to provide comparative data. PEFA assessments can be expensive—costing over $100,000 per country in some cases—and the results are not always made public, depending on the preferences of the countries themselves. Nevertheless, they typically provide a fairly robust and comprehensive assessment of where PFM systems are working effectively and where they are not.

In a similar fashion, the UNCAC Treaty—which currently has over 140 signatories—mandates a multi-stage review process of where its members

stand along a broad number of corruption-related dimensions. The treaty has four chapters which focus on preventive measures, criminalisation and law enforcement, international cooperation and asset recovery. It covers everything from the legal and institutional framework for combatting corruption to collaboration in law enforcement. It has a three-stage review process in which an initial self-assessment is followed by an independent peer-review process. The strengths of the UNCAC review process are its comprehensiveness and the fact that it covers both the preventive and enforcement dimensions of the anticorruption agenda. The downside is that the full review process can be lengthy and stretch out over several years and is typically not repeated. Also, unlike PEFA, the UNCAC review process has so far not been used for generating comparative data and benchmarking countries against each other.

Over the past five years, there has been increasing interest within the donor community on the importance of combatting illicit financial flows. Much of this dialogue has taken place in the context of financing for the Sustainable Development Goals (SDGs), where issues of domestic revenue mobilisation are looming increasingly large. To date, much of the analytical work surrounding illicit financial flows has been spearheaded by the Washington-based NGO Global Financial Integrity (GFI), which produces annual financial estimates regarding the size of these flows. Their most recent estimate for 2013 is staggering —nearly $1.1 trillion. According to GFI calculations, more resources are flowing out of Africa than are flowing into the continent, with the net beneficiaries being banks in the USA, the UK, Switzerland and various off-shore tax havens (Illicit Financial Flows 2017). GFI seeks to break down these flows by type (faulty trade invoicing, money laundering, etc.) and country.

The study of international flows remains relatively underdeveloped, and a number of scholars have questioned GFI's methodology and whether the size of these figures is accurate (Reuter 2012). It may not adequately differentiate between legitimate strategies for minimising the tax burden, such as transfer pricing, and illicit practices such as money laundering and the creation of shell companies. Yet even if the figures are a fraction of GFI's estimates, the problem remains serious. By way of comparison, the total aid flow from OECD countries to the developing world for 2014 was $137 billion.

Approaches such as PEFA, UNCAC assessments and work on illicit financial flows play a vital role in illuminating certain dimensions of the broader anticorruption effort and providing a much more detailed picture of how improvements can be made. They are therefore invaluable to a number of important constituencies, such as policymakers, donors, academics

and representatives from civil society. However, their technical nature and complexity often do not lend themselves well to broader public debate—a reality further complicated by the fact that, in some cases, the findings may not be made available to the general public.

Leading Global Indices: Transparency

An important sub-set of the broader good governance and anticorruption debate involves the promotion of transparency and the right to information. In 2015, this effort was embedded in SDG Goal No. 16 on promoting peaceful and inclusive societies for sustainable development and building effective, accountable and inclusive institutions. Beneath this goal, SDG Target 16.6 explicitly embraces the need to develop "effective, accountable and transparent institutions at all levels" (UNDP 2017).

Development organisations such as the World Bank have been strongly supportive of efforts to enhance transparency for a variety of reasons. As Fig. 7.1 indicates, although causality can be difficult to attribute, there is a strong correlation between indices of government effectiveness and those for open government. Studies have noted evidence that, in some contexts, greater fiscal transparency can result in lower government borrowing costs. There is also growing empirical evidence that greater transparency and public awareness can improve resource flows to front-line organisations, facilitate social accountability and the demand for better services, and reduce political corruption and strengthen accountability to citizens (Glennerster and Shin 2008;

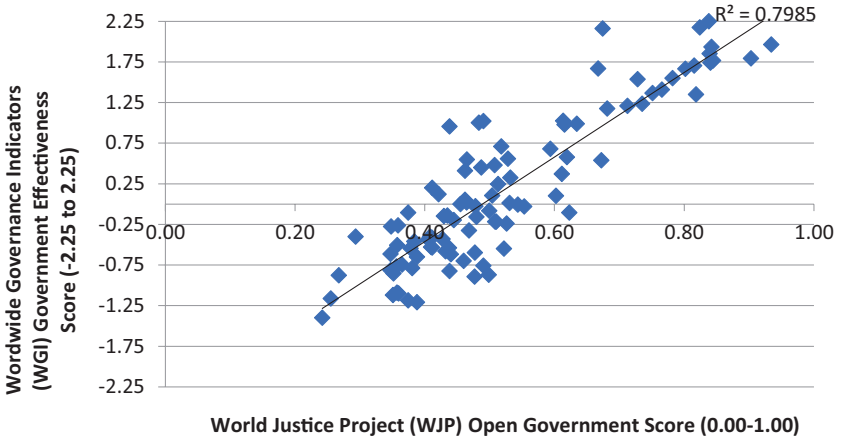

Fig. 7.1 Government effectiveness and open government, 2011

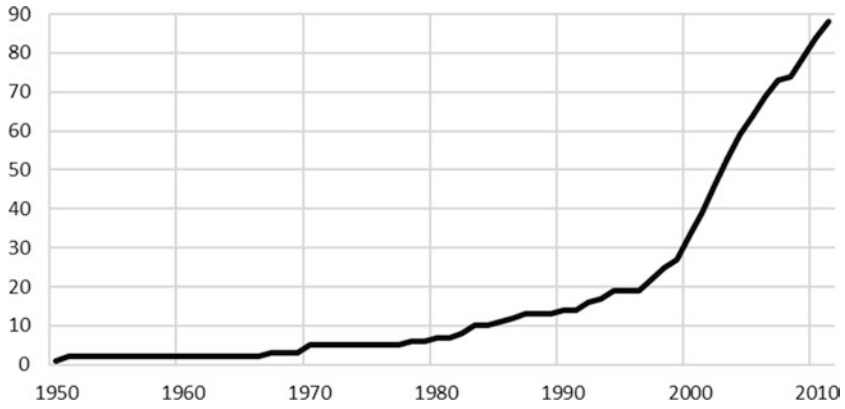

Fig. 7.2 Global increase in right to information legislation. Source: World Bank staff calculations

Reinikka and Svensson 2005). As research in this area evolves, it is becoming increasingly clear that transparency alone is not sufficient. The wider context matters too, as do factors such as the engagement of civil society organisations and the prevalence of corruption throughout society. There is also some evidence that enhanced transparency and public participation can shift corruption to areas where scrutiny is less pronounced (Etter 2014; Bauhr and Grimes 2013; Olken 2007). Such findings serve more as qualifications than as refutations of the broader contention that transparency serves as an important dimension in both combatting corruption and enhancing government effectiveness.

The global movement towards increased transparency is one of the defining governance trends of the past two decades. Figure 7.2 provides information on the growth of right to information legislation—a key indicator of the broader global transparency movements. This agenda picked up considerable momentum in the mid-1990s and has continued its rapid growth since. As of late 2016, 114 countries on every continent have implemented some variant of RTI legislation. Other important dimensions of the global transparency movement include a push towards open data and a move towards fiscal transparency.

It spite of the rapid growth of the movement towards increased transparency and the right to information, the development of indices to measure these phenomena has lagged considerably behind those relating to anticorruption. There are currently two indices for transparency that are in widespread circulation. The first is the Global Right to Information Index, which is published jointly by Access Info Europe and the Centre for Law and

Democracy. Initiated in 2010, the rating seeks to assess the overall strength of RTI legislation. Countries are ranked using 61 indicators covering seven general categories, with a maximum of 150 points.

The second major index of transparency is the Open Budget Survey produced by the International Budget Partnership. The 2016 Open Budget Survey includes data on 102 countries and covers the transparency of eight dimensions of the broader budget cycle, ranging from the pre-budget statement and the executive's budget proposal to the enacted budget and the citizens' budget, to mid-year reviews and the final audit report.

Beyond these measures, the criteria for joining the Open Government Partnership could constitute a third de facto set of rankings. Countries interested in OGP membership need to meet four sets of criteria: (1) fiscal transparency, (2) access to information, (3) asset disclosure for public officials and (4) citizen engagement. They are scored on a scale of 1–4 points for each dimension with a total of 16 points, and they need a score of 75 per cent or better to qualify to join the OGP. As of 2016, 69 countries have joined.

While these efforts mark important steps forward, more work remains to be done. This is particularly true with regard to monitoring other dimensions of transparency, such as open data, and in developing more nuanced assessments regarding the actual implementation of RTI regimes in addition to the quality of RTI legislation. Some interesting efforts have been initiated in this direction recently, and it is possible that momentum will increase as a result of SDG Target 16.6 and the need to measure compliance with that objective.

Conclusion

As the discussion above underscores, the effort to create indices that accurately capture and reflect complex phenomena such as corruption and transparency has made considerable progress in the past two decades. The field continues to evolve in ways that are both interesting and constructive. However, for all of the reasons outlined above, the underlying complexity of the issues under consideration is likely to continue to defeat a precise understanding of where countries fall in any sort of global ranking. To draw upon CPI's latest 2015 index, for example, few would dispute that Denmark, Sweden and New Zealand have succeed in advancing the cause of anticorruption throughout their governments in ways that Sudan, Libya and Pakistan have not. One can probably draw distinctions with some degree of comfort between countries that are relatively clean, those that struggle with modest

amounts of corruption, those falling somewhere in the mid-range and those that are confronted with serious challenges. But more nuanced gradations, such as those between Thailand (45), Croatia (50) and Kuwait (55), are unlikely to be large enough to be meaningful.

However, progress is likely to occur along several dimensions. First, one can hope that the robustness of cross-country comparisons will marginally increase over time. Initial assessments tended to indicate gaps between perception-based assessments and other metrics. More recent research appears to be affirming that under an increasing number of circumstances, perception-based assessments are relatively consistent with expert views and can indeed serve as a proxy for how states rank against each other (Charron 2016).[2] Used with caution and humility, indices such as the CPI and the WGIs can provide a rapid cost-effective first approximation of the nature of corruption-related challenges confronting a given polity. They are useful to frame the debate and get the broader discussion going regarding the urgency and severity of a problem. They can also serve as a call to action, although what type of action remains very much an open question, as noted above.

Second, it is encouraging to see more detailed and systematic treatments along the lines of the UNCAC, PEFA or the (former) Global Integrity assessments, which will ultimately help to provide a more detailed menu of actionable reforms that can help guide policymakers. These assessments do not readily lend themselves to cross-country comparisons, although some such as PEFA are occasionally used in this context. They are effective at reviewing the efficacy of the underlying legal and institutional framework shaping a given country's anticorruption effort—at least for those dimensions falling within their purview—and at pointing the way to where reforms are needed.

These developments raise the broader question of how accurate our indices need to be. To return to the CPI example cited above, most reasonable observers would not impart much significance to the distinctions between countries such as Thailand, Croatia and Kuwait beyond noting that they fall in the mid-range or perhaps the upper mid-range of countries in terms of perceptions of corruption. For most countries, their relative standing vis-à-vis regional comparators is often a much more important and meaningful benchmark. In this context, the average Kuwaiti, for example, would see himself or herself as falling behind Qatar (22) and the UAE (23) and generally in the company of Jordan (45), Bahrain (48) and Saudi Arabia (50). Kuwait's situation would be seen as better than that of most middle-tier countries in MENA—Egypt and Morocco are at 88, for example—and well ahead of the laggards (Iraq and Libya come in at 161). This could be a cause for rejoicing or lament. But if it does serve as a call for action, one would then move on to a host of more

proximate assessments and metrics that provide greater nuance and detail as to the nature of the problem at stake and what Kuwait's next steps should be.

Our global indices are therefore likely to serve as navigation tools, not unlike our contemporary GPS systems. One can trust them to provide general guidance from point A to B, but not to park one's car. Actual decision-making, whether in business, government or among donor agencies, such as the World Bank, requires a granularity of information that is simply not present in broader global indices. In fact, even the more detailed cross-country technical assessments are ultimately inadequate. Just because a law is lacking or an institution needs strengthening in certain well-defined areas does not necessarily mean that it will happen. A variety of idiosyncratic factors—political windows of opportunity, bureaucratic support and opposition, the talent and/or enthusiasm of a given minister, and so on—will always require careful probing and analysis. Neither indices nor assessments will yield any information on the critical question of how to implement reforms, and tactics matter.

Nevertheless, our tools have reached a level of sophistication at which, taken collectively and used judiciously, they can help assess the magnitude of problems and suggest the areas where greater focus is needed. They can, on occasion, even serve as a spur towards action. These are not trivial accomplishments and are a tribute to the progress that has been made over the past two decades in advancing our understanding of this subject.

Notes

1. TI changed its methodology in that year to facilitate improved inter-temporal comparisons, with the understanding that changes after 2012 could not be compared with scores from previous years.
2. Charron also notes that perception indicators, in spite of showing strong validity (at least within the sample of European states considered in his analysis), are best used as ordinal measures to compare how states or regions rank relative to one another rather than being used as hard "benchmarks" to assess actual levels of corruption.

References

Ahmed, A. (2014, December 3). Pakistan's Ranking in Corruption Index Improves. *Dawn Newspaper*. Retrieved from http://www.dawn.com/news/1148425

Arndt, C., & Oman, C. (2006). *Uses and Abuses of Governance Indicators*. Paris: OECD.

Bauhr, M., & Grimes, M. (2013). Indignation or Resignation: The Implications of Transparency for Societal Accountability. *Governance, 27*(2), 291–320.

Charron, N. (2016). Do Corruption Measures Have a Perception Problem? Assessing the Relationship Between Experiences and Perceptions of Corruption Among Citizens and Experts. *European Political Science Review, 8*(1), 147–171.

Donachev, D., & Ujhelyi, G. (2014). What Do Corruption Indices Measure? *Economics and Politics, 26*(2), 309–331.

Etter, L. (2014, February 20–21). *Can Transparency Reduce Corruption? Evidence from Firms in Peru and Mali on the Impact of the Extractive Industries Transparency Initiative on Corruption*. (Prepared for the Doing Business Conference at Georgetown University). Washington, DC: Georgetown University. Retrieved from https://repository.library.georgetown.edu/handle/10822/557839

Glennerster, R., & Shin, Y. (2008). *Does Transparency Pay?* (International Monetary Fund Staff Papers, 55/1, 183–209). Washington, DC: International Monetary Fund.

Global Financial Integrity. (2017). *Illicit Financial Flows*. Washington, DC: Global Financial Integrity. Retrieved from http://www.gfintegrity.org/issue/illicit-financial-flows/.

Governance and Social Development Research Centre (GSDRC). (2010). *Helpdesk Research Report: Critique of Governance Assessment Applications*. London: GSDRC. Retrieved from http://www.gsdrc.org/docs/open/hd699.pdf.

Hough, D. (2016, January 27). Here's This Year's (Flawed) Corruption Index. Those Flaws are Useful. *The Washington Post*. Retrieved from https://www.washingtonpost.com/news/monkey-cage/wp/2016/01/27/how-do-you-measure-corruption-transparency-international-does-its-best-and-thats-useful/?utm_term=.2b6a199a103b

Kaufmann, D., & Kraay, A. (2007, January 11). *On Measuring Governance: Framing Issues for Debate*. (Issues Paper for the Roundtable on Measuring Governance Hosted by the World Bank Institute and the Development Economics Vice-Presidency of The World Bank). Washington, DC: World Bank. Retrieved from http://www1.worldbank.org/publicsector/anticorrupt/corecourse2007/OnMeasuringGovernance.pdf

MacMullen, R. (1988). *Corruption and the Decline of Rome*. New Haven: Yale University Press.

Olken, B. (2007). Monitoring Corruption: Evidence from a Field Experiment in Indonesia. *Journal of Political Economy, 115*(2), 200–249.

Pritchett, L., Woolcook, M., & Andrews, M. (2010). *Capability Traps? The Mechanisms of Persistent Implementation Failure*. Harvard University: Center for Global Development. Retrieved from https://www.hks.harvard.edu/fs/lpritch/Governance/capability_traps(june2010).pdf.

Reinikka, R., & Svensson, J. (2005). Fighting Corruption to Improve Schooling: Evidence from a Newspaper Campaign in Uganda. *Journal of the European Economic Association, 3*(2–3), 259–267.

Reuter, P. (Ed.). (2012). *Draining Development? Controlling Flows of Illicit Funds from Developing Countries*. Washington, DC: World Bank Publications.

Rijkers, B., Freund, C. L., & Nucifora, A. (2014). *All in the Family: State Capture in Tunisia* (World Bank Policy Research Working Paper No. 6810). Washington, DC: World Bank. Retrieved from http://documents.worldbank.org/curated/en/440461468173649062/All-in-the-family-state-capture-in-Tunisia

Thomas, M. A. (2010). What Do the Worldwide Governance Indicators Measure? *The European Journal of Development Research, 22*(1), 31–54.

Transparency International. (2012). *Corruption Perceptions Index 2012: An Updated Methodology* [Press Release]. Berlin: Transparency International. Retrieved from http://www.transparency.org/files/content/pressrelease/2012_CPIUpdatedMethodology_EMBARGO_EN.pdf

Transparency International Ukraine. (2012). *Corruption Perceptions Index 2012: Techincal Methodology Note*. Kyiv: Transparency International Ukraine. Retrieved from http://old.ti-ukraine.org/en/content/2409.html.

United Nations Development Programme (UNDP). (2017). *Sustainable Development Goals*. New York: UNDP. Retrieved from http://www.undp.org/content/undp/en/home/sustainable-development-goals.html.

Williams, S. (n.d.). Review of Corruption and the Decline of Rome. *History Today*. Retrieved from http://www.historytoday.com/stephen-williams/corruption-and-decline-rome

World Justice Project. (2017). (Website). Retrieved from http://worldjusticeproject.org/

Part II

Making Measures

8

Understanding Governance and Corruption Using Survey Data: A Novel Approach and Its Applications in Policy and Research

Francesca Recanatini

Introduction

On 1 October 1996, James Wolfensohn, then President of the World Bank Group, delivered his address to the annual meetings of the World Bank and the International Monetary Fund, and for the first time in the history of these organisations, he explicitly mentioned the "cancer of corruption":

> If the new compact is to succeed, we must tackle the issue of economic and financial efficiency. But we also need to address transparency, accountability, and institutional capacity. And let's not mince words: we need to deal with the cancer of corruption. (International Monetary Fund 1996, p. 27)

Twenty years have passed since then and both development practitioners and researchers find themselves in a very different place today than our colleagues in 1996. Increasingly, researchers have focused on the link between

Prepared for the Workshop on Global Governance by Indicators: Measuring Governance and Stateness, Florence, 13–14 November 2014. This work would not have been possible without the cooperation of thousands of survey respondents who selflessly shared their experiences. The views expressed in the paper are those of the author and should not be attributed to the World Bank, its executive board, or its management.

F. Recanatini (✉)
World Bank, Washington, DC, USA

D.V. Malito et al. (eds.), *The Palgrave Handbook of Indicators in Global Governance*,
https://doi.org/10.1007/978-3-319-62707-6_8

poor governance, corruption, and economic growth.[1] Cross-country evidence has shown that poor governance and corruption can be harmful to citizens' standards of living and income distribution—reducing literacy and per capita income while increasing infant mortality. Furthermore, poor governance distorts public expenditure and increases poverty, thus reducing investment efficiency (see Kaufmann 2000; Knack and Anderson 1999; Gupta et al. 1998; Mauro 1995; Tanzi and Davoodi 1997; World Bank 2004; Svensson 2005; Bardhan 1997; Broadman and Recanatini 2002; Fisman and Gatti 2002; Friedman et al. 2000; Jain 2001; Johnson et al. 2000; Treisman 2000). In particular, by promoting misallocation of resources and poor accountability, bad governance may raise the costs (both official and unofficial) of public services and limit users' access to them. Users with lower incomes or firms of different sizes may pay more or less in terms of bribes than others to obtain the same public service (Hunt and Laszlo 2005; Svensson 2003).

However, this wealth of cross-country knowledge and information has not been associated with significant progresses in improving governance and government institutions in countries. Moreover, new data has shown that the quality of governance and institutions can differ significantly within the same country (Charron 2013). This has led to the realisation of the complexity of the governance challenge and the limits of the existing approaches and cross-country indicators in policy-making and reform implementation. In particular, practitioners have realised the importance of better understanding and evaluating governance within a country and at the agency level. A study of the business environment across Russian regions (Kisunko et al. 2013) is an example of the move towards more disaggregated data and analysis.

This chapter builds on a decade of work carried out by the World Bank on governance and anti-corruption (AC) reforms and shares an alternative approach to evaluating governance within a country for policy design and implementation. Section "Improving Governance and Tackling Corruption: A Few Initial Lessons" offers an overview of the initial lessons on governance and AC reforms and sets the stage for the methodological approach described in section "Gathering Disaggregated Information and Data: An Alternative Methodology". The section "Building Agency-Specific Indicators" introduces the new agency-specific indicators of governance, highlighting their applicability and limits. The section "Applications" discusses the applications of these indicators, and the section "Conclusions" offers some concluding remarks and suggests areas for further work.

Improving Governance and Tackling Corruption: A Few Initial Lessons

Governance and corruption are multi-dimensional phenomena that affect different sectors and groups of citizens in a country. Their complexity and multi-sectoral breadth have led to different methods of defining and measuring them and, in turn, to different approaches and policy responses to address them. Four specific features have emerged as common to all countries when thinking about governance and corruption:

- The heterogeneity of these phenomena;
- The importance of implementing reforms that focus on changing incentives and behaviours rather than laws and regulations;
- The need for a significant reallocation of powers and rents to improve governance and address corruption; and
- The importance of sustaining reform efforts over time.

Together, these four features pose a challenge to policy-makers and researchers when attempting to understand and address poor governance and corruption. The first two features can have an impact on the type of information needed to evaluate these phenomena, to understand their potential causes, and to identify effective policy tools. The last two features, instead, can affect the process of designing and implementing reforms and their success. I will discuss each feature separately.

First, in practice, governance and corruption are *heterogeneous,* that is, they can take different forms within the same country and across different countries. Box 8.1 offers examples of some different forms of corruption World Bank practitioners have come across in their work. It is clear that cross-country indicators using a broad definition of these problems provide an imperfect measure of these variables. This can have an impact both on the efforts of policy-makers and the work of researchers.

Box 8.1 Typologies of Corruption

- *Corruption in public procurement*: use of bribes, gifts, and/or favours to alter the public procurement process;
- *Corruption in budget management*: use of bribes, gifts, and/or favours to influence budget management decisions and divert funds;
- *Corruption in personnel management*: use of bribes, gifts, and/or favours to affect personnel management decisions;
- *Legal and regulatory corruption*: use of bribes, gifts, and/or favours to alter regulatory and legal decisions;
- *Administrative corruption* (*or petty corruption)*: use of bribes, gifts, and/or favours to obtain or hasten the provision of public services.

Source: Own compilation

In particular, policy-makers need to understand the specific characteristics of the problem they face and then use the most appropriate policy instruments to address that particular problem. In addition, policy-makers need to consider the *heterogeneity* of the policy environment in which a policy will be implemented: each country faces a different set of political, institutional, and cultural conditions and constraints. Countries are also endowed with different levels of skills and resources that can be used to address these issues. Moreover, different actors and stakeholders (local, national, and international) can play different roles in support of, or against, AC reforms. These country-specific differences need to be taken into account when designing and implementing successful AC policies.

This heterogeneity of conditions and manifestations of the phenomena can also affect the ability of researchers to understand the potential causes behind poor governance and corruption and to identify the most effective policy tools if they are using aggregated measures for their research. The first generation of research and analysis described in the introduction to this chapter has been helpful in providing evidence of the links between governance, corruption, and growth, but researchers are now focusing on a new set of questions and are delving more into specific elements and factors that affect governance and can lead to corruption. This requires information and data that are much more specific and detailed than the existing cross-country indicators can provide.

Second, along with a push for more disaggregated information, practitioners and researchers are increasingly realising the role of incentives and behaviour, rather than that of laws and regulations when it comes to improving governance and reducing the risk of corruption. Laws and regulations are necessary to establish and define transactions and exchanges among citizens and between citizens and government institutions, but they are not sufficient to guarantee that these transactions will happen in the way specified by the law. In practice, laws and regulations will be implemented differently and to different degrees depending on the system of incentives present in the country concerned.

Researchers and practitioners have realised the existence of this "gap" between laws and their actual implementation and the importance of focusing on the system of incentives driving exchanges in a country. In particular, researchers now realise that addressing the issue of poor governance and corruption entails a change in the way in which citizens, business people, and public officials interact with each other. This requires an understanding of the system of incentives that regulates the relationships between citizens and the government and a set of policies that can change it. The first wave of AC

reforms often focused on the introduction of new laws to curb corruption. These laws were only partly built on knowledge of the system of incentives that already existed in countries and with limited understanding of the resources and steps needed to implement them. As a result, these laws were at best only partially implemented and led to very small changes in the behaviour of citizens, business people, and public officials.

This has forced practitioners and researchers to focus more on the system of incentives that regulate interactions among different economic agents and to understand more precisely what is necessary to support the implementation and use of a new law. In turn, this has created the need for data that captures actual behaviour and the degree of implementation of existing laws and regulations.

Third, governance and corruption are not only heterogeneous phenomena that require a focus on behaviour and incentives. They also require a reallocation of roles and responsibility within a country, that is, a reallocation of powers. These issues may be even more severe for certain types of corruption, particularly where rents are large and interest in maintaining the status quo is strong.[2] As a result, measures to reduce the incidence of petty corruption may receive wider and stronger political support than measures that address other types of corruption (such as the diversion of public funds or corruption in public procurement), since control of petty pay-offs will only have a limited impact on the distribution of the most significant rents and power in a country. Thus, when considering AC reforms, strong leadership and a continuous commitment from the top are the first necessary ingredients as these send a clear signal against the maintenance of the status quo. The second necessary ingredient is the creation of channels that allows citizens and business people to support and sustain reform efforts from the bottom. Both these ingredients should be supported by a deep understanding of the political landscape of the country in question. Insight into the impacts of different stakeholders on the implementation of a reform should translate into a mix of policy measures that can help mitigate the risks of partial implementation and reversal of policies by those opposed to change.

The heterogeneity of governance and corruption and the importance of focusing on changes in behaviour highlight the need to adopt a medium/long-term approach. This points to a fourth and final feature of these phenomena. Because profound institutional changes are needed in order to successfully improve governance and address corruption, when they address political-economy issues, reform packages need to include mechanisms that foster the sustainability of the reform process. Concretely, policy-makers should involve citizens when designing reforms and include in the final reform

package a set of tools to measure progress achieved on the ground and to communicate progress to citizens. A regular and continuous dialogue with citizens and supporters of the reforms on concrete issues can help generate ownership of the reform process and can foster support (see Kpundeh and Stapenhurst 1999; Johnston and Kpundeh 2004, for an extensive discussion of this issue). Building on these considerations, policy-makers should complement short-term policy measures that can lead to immediate observable changes with deeper medium-term reforms that can permanently change the allocation of powers and rents. Only in this way—mixing "quick-wins" with medium-term measures and regular communication about progress—can policy-makers guarantee the support of citizens and less resistance from interest groups throughout the implementation of reforms.

In sum, after more than two decades of work and research on governance and corruption, four features appear to undermine our ability to identify the most appropriate policy tools and implement effective policy reforms: the heterogeneity of these phenomena and the limited existence of disaggregated data able to capture this variation; the need to focus on behaviours rather than rules; the importance of changing the system of incentives; and the need to support reform efforts over the medium and long term. A participatory approach to policy formulation and implementation can help mitigate difficulties with the latter two issues, while more disaggregated data focused on experience is needed to address the former two challenges. The next section presents an approach focused on micro-experiential data that can begin to address these issues.

Gathering Disaggregated Information and Data: An Alternative Methodology

The previous considerations—heterogeneity of the phenomena, the role of incentives, the need for a reallocation of power, and the sustainability of reforms—and the limitations of existing aggregated indicators have pushed practitioners and researchers to develop alternative approaches to the design of policy reforms. In particular, a team of World Bank practitioners have introduced a new approach focused on micro-level data collected from different stakeholders.[3] The instruments used in this approach are detailed country-specific surveys of households, business people, and public officials, called Governance and Anti-corruption (GAC) Diagnostics. These tools provide a basis for technical discussions by policy-makers and civil society on policy formulation and complement other sources of information.

The GAC diagnostic surveys were originally developed to respond to specific country demands for the design of a governance strategy. Since their first implementation in 1998 in Latvia, Georgia, and Albania, they have been adapted and implemented in about 18 countries.

This approach comprises the following key features:

- Multi-pronged surveys of users of public services/households, business people, and public officials, permitting triangulation of responses;
- The collection of experience-based (vs. "opinions") types of information; and
- Participatory implementation of the data analysis and of the drafting of the strategy (to include government representatives, civil society, media, parliamentarians, business people and the donor community, etc.).

The objective of this approach is to improve the quality of public services by enhancing the governance systems responsible for personnel and resource management, access to services, and citizen feedback mechanisms. The approach has allowed practitioners to:

- Unbundle the problems and manifestations of governance and corruption;
- Explore the channels through which these issues manifest themselves;
- Uncover weak and strong institutions to define the potential sources of the problem;
- Evaluate the impact and costs of poor governance and corruption for different stakeholders;
- Identify concrete and measurable ways to reduce these costs through targeted reforms.

The template developed for the implementation of this approach includes the following phases:

- ***A preparatory phase*** (Phase 0) which is used to identify and recruit the project team and develop a detailed work programme.
- ***A "partnership-promoting" phase*** (Phase 1), which includes the implementation of a series of focus groups (with relevant local counterparts) to facilitate the coordination of the different national stakeholders involved and adapt existing tools to the challenges faced by the specific country.
- ***A development phase*** (Phase 2) to integrate the outcome of the focus groups, to revise the existing templates for the diagnostic tools, and to train local staff for the necessary fieldwork.

- ***A fieldwork phase*** (Phase 3) to implement the surveys, using the tools developed in Phase 2. This phase includes sample design, fieldwork, and data coding and is carried out jointly by local and external experts.
- ***An analytical phase*** (Phase 4) to analyse the data collected in Phase 3 and to begin to identify potential areas for policy actions. The analysis is carried out jointly by the Bank team and the local team.
- ***A validation and dissemination phase*** (Phase 5) to validate the findings through focus group discussions and share the results of the analysis completed in Phase 4 with local and central governments, citizens, the media, and research agencies.

The Governance Anti-Corruption Diagnostic survey portal offers an overview of the tools, the manuals, and the methodology in several different languages (English, Spanish, French, Arabic, and Kurdish). It also provides access to the complete reports produced and published by various countries.

The design and implementation of the surveys supporting this approach have provided a wealth of information and data on governance systems and corruption, which allow researchers to begin to untangle the possible factors that can contribute to better systems. In addition, these surveys have helped establish initial benchmarks for the performance of public sector institutions in countries to allow policy-makers to monitor quality of governance and public sector performance on a regular basis both at the national and at the sub-national levels. Finally, the reform plans developed using this comprehensive and participatory approach have a better chance of being implemented by governments.[4]

This approach begins to address the four challenges discussed in the previous section in several ways. The surveys help inform the reform process by providing detailed and factual information about the strengths and weaknesses of the system analysed—information that is triangulated and validated by different stakeholders. The participatory approach—used to finalise the instruments for the data collection, to validate the results, and to draft policy recommendations—creates the foundation for greater ownership and sustainability of the reform process at the country level. Finally, the explicit political commitment of high-level government officials creates space for a constructive policy debate and for the implementation of difficult reforms. Together, these three elements begin to address the challenges policy-makers still face when attempting to measure governance and corruption for policy reforms, and they offer an approach that promotes capacity building at the country level.

Building Agency-Specific Indicators

The country-specific survey data described in the previous section provide a wide range of information about governance, corruption, and service delivery across key pre-selected institutions within a country based on responses from citizens, business people, and public officials. These data can be analysed to evaluate the quality of governance across agencies, to identify both weak institutions (in need of reform) and strong institutions (examples of good governance) and to unbundle corruption in its various forms—such as administrative corruption, state capture, corruption related to procurement, theft of goods and public resources, and purchase of licences and regulations.

Simple analysis of the survey data can make it difficult to identify specific factors and variables that may contribute to higher or lower levels of governance and corruption. Therefore, a World Bank team led by Daniel Kaufmann and Francesca Recanatini decided to create a set of governance indicators using responses from the public officials. The purpose of this effort was to make the abstract concept of governance more concrete and operational. Rather than focusing on the individual questions, the proposed approach groups' answers to questions on similar issues to provide an assessment of each specific dimension of governance (e.g. transparency in the management of the budget). The starting point for the grouping of the individual answers was the approach introduced by Kaufmann et al. (2002) for analysis of public agencies in Bolivia. Using this approach, we constructed indicators on different aspects of governance such as transparency, meritocracy, openness, and accountability. The data from the survey of public officials allowed us to create more precise measures of these aspects—such as, for example, the existence of a merit-based system for hiring civil servants—and for this to be done at the agency and regional levels rather than the country level. This significantly increases our ability to identify governance challenges and to spot areas of excellence.

To create these indicators, we first rescaled the individual responses from the public officials and synthesised them into agency indices using a factor analysis technique. In simple terms, this technique extracts common information from different sources (in this case, the different answers provided by civil servants) and uses this common information to construct new variables. In particular, all the questions about individual perceptions, which may be scaled from 1 to 7 (or 1 to 5) in the original survey, are rescaled from 0 to 100 in order to facilitate interpretation in percentage terms (with 0

always meaning the lowest level of quality of governance, corruption, access, or service performance). We then constructed several variables intended to capture different aspects of governance quality, the presence of corruption within institutions, and access to and availability of public services provided to citizens.

Table 8.1 provides definitions of the governance indicators constructed in this way and Table 8.2 describes the various corruption measures we have developed. The individual questions used for the construction of each indicator are presented in Annex 8.1 along with some country illustrations.

Kaufmann et al. (2002) approach allows respondent bias and measurement error due to individual differences in perceptions to be minimised since it uses the common information that exists across answers through the application of factor analysis.[5] In addition, it is possible to separate agencies located in different cities and introduce them into the sample as different units of observation, one per city. This attempts to capture the idea that the quality of governance and corrupt practices within the same agency may differ in a significant way depending on the geographical location of the agency, which leads to a more accurate picture of the quality of governance and a more significant analysis.

In practice, this way of grouping responses from public officials allows us to obtain a better picture of the specific institutional and governance challenges faced in a country. Consider, for example, the data for Peru (World Bank 2001) summarised in Table 8.3. The governance indicators are presented for the whole country and for each of its four regions. It is immediately apparent that the extent of corruption varies not only across regions but also across types of corruption. Corruption regarding contracts is reported to be the most serious form of corruption in the Selva region, while in the Lima region, budgetary corruption appears to be more severe. The indicators also highlight significant regional differences in terms of governance dimensions.

The responses from public officials can also be grouped by the agency where the survey respondents work, rather than by the region where they are located. The indicators constructed in this way—such as, for example, those presented in Table 8.6 in Annex 8.1 for World Bank (2003)—offer a view of which public agencies are performing better when it comes to managing human resources, budgeting or delivering services. This can help policy-makers to focus their efforts on the agencies where the needs are greater, at the same time using the experience of the well-performing public agencies operating in the same country and under the same regulations.

Table 8.1 Definitions of governance indicators

Governance variable	Definition
Access to the poor	Index of the accessibility of the public service provided by the institution for poor citizens. High numbers imply high levels of access.
Service performance	Index indicating the extent to which the service provided by the institution is of high quality, fully satisfactory to the user, and offered at relatively low cost.
Audit mechanisms	Percentage of cases where decisions on personnel, budget, procurement, and service management are subject to external and/or internal audits. In the empirical analysis, this is subdivided into audits of service management, personnel management, procurement management, and budget management.
Enforcement of rules	Frequency with which rules/guidelines/regulations on personnel, budget, procurement, and service management are enforced.
Supervision of rules	Frequency with which rules/guidelines/regulations on personnel, budget, procurement, and service management are supervised.
Clarity of rules	This index combines the frequency with which rules/guidelines/regulations on personnel, budget, procurement, and service management are well supervised and strictly enforced. It is subdivided into four categories: supervision of personnel rules, budget rules, procurement rules, and service rules.
Meritocracy	Frequency with which decisions on personnel management issues are based on professional experience, merit, performance, or education levels.
Mission	Index determining the degree of understanding of the agency's purpose and their own tasks and responsibilities by public officials within the institution.
Openness	Frequency with which rules/guidelines/regulations on personnel, budget, procurement, and service management are announced and opened to the inside of the institution. It is subdivided into four categories: openness of personnel decisions, budget decisions, procurement decisions, and service decisions.
Politicisation	Percentage of cases where decisions on personnel, budget, procurement, and service management are subject to political interference. In the empirical analysis, it is subdivided into politicisation in service management, personnel management, procurement management, and budget management.
Quality of rules	Frequency with which rules/guidelines/regulations on personnel, budget, procurement, and service management are formally written, simple and clear, and do not add too many administrative steps. It is subdivided into four categories: quality of personnel, budget, procurement, and service management rules.

(*continued*)

Table 8.1 (continued)

Governance variable	Definition
Resources	Percentage of cases where the agency's physical, financial, and human capital resources are adequate.
Transparency	Percentage of cases where decisions on personnel, budget, procurement, and service management are made transparently.
Voice	Index representing the existence of consumer feedback and complaint mechanisms.
Wage Satisfaction	Percentage of employees who are very satisfied or somewhat satisfied with their wages and benefits.
Rotation	Index indicating the frequency of the agency rotation rate between the public and private sectors.

Source: Own compilation

Table 8.2 Definitions of corruption indices

Index	Definition
Overall corruption index	Corruption index representing the average of five standardised corruption indicators: bribery over regulatory/legal decisions, bribery over public contracts, bribery to obtain public services, corruption in personnel management, and corruption in budget management.
Sub-indices of corruption analysed:	
Corruption in personnel management	Corruption index representing the percentage of cases where decisions on personnel management are based on unofficial payments (job purchase).
Corruption in personnel management(2)	Corruption index representing the extent of the practice of "purchasing jobs" in the institution.
Corruption in budget management	Corruption index representing the frequency of irregularities/diversion of funds or any other type of budget abuse within the institution.
Corruption in contracts	Corruption index representing the frequency of bribes in the contracting process within the institution.
State capture	Corruption index representing the extent of bribes to alter regulations and legal decisions within the institution.
Administrative corruption	Corruption index representing the frequency of bribes to obtain public services within the institution.

Source: Own compilation

Applications

The previous sections have introduced the approach and the governance and corruption indicators we have developed. These indicators are clearly quite useful from a policy point of view since they more accurately help identification of

Table 8.3 Governance and corruption indicators by region, Peru

Governance and corruption indicators	Lima	Sierra	Selva	Resto de Costa	Whole country
Overall corruption	32	25	29	29	30
Corruption in personnel	24	20	20	23	22
Corruption in budget	35	25	24	24	29
State capture	23	19	23	22	22
Corruption in contracts	24	17	26	22	23
Administrative corruption	24	20	24	23	23
Perceived corruption	44	39	42	43	43
Accessibility for poor	67	75	71	70	70
Audit mechanisms	60	68	63	59	62
Audit mechanisms of personnel decisions	60	70	60	58	62
Audit mechanisms of budget decisions	62	69	66	63	64
Audit mechanisms of public contracts	59	65	61	58	60
Enforcement of rules	47	51	47	46	48
Supervision of rules	70	76	72	71	72
Clearness of rules	55	63	59	59	58
Clearness of personnel rules	53	66	57	57	57
Clearness of budget rules	58	60	60	59	59
Clearness of public contracts rules	56	63	60	59	59
Quality of rules	67	76	68	69	69
Quality of personnel rules	69	76	66	70	70
Quality of budget rules	67	76	68	69	69
Politicisation	36	31	32	34	34
Politicisation in personnel decisions	37	28	30	34	33
Politicisation in budget decisions	36	34	33	34	34
Politicisation in public contracts	34	31	33	36	34
Resources	55	61	60	53	57
Transparency	56	65	61	60	59
Transparency in personnel decisions	51	63	57	57	55
Transparency in budget decisions	59	67	63	63	62
Transparency in public contracts	58	66	62	60	61
Openness	53	61	58	58	57
Openness in personnel decisions	55	69	57	58	59
Openness in budget decisions	54	55	58	56	55
Openness of public contracts decisions	53	61	59	59	57
Mission	78	85	80	81	80
Service performance	64	70	65	64	65
Meritocracy	63	73	66	64	66
Citizen voice	68	73	68	70	69
Wage satisfaction	37	40	38	36	38

Source: World Bank 2001

The indicators above take values between 0 and 100. To interpret them, it should be kept in mind that:

The higher the value of the corruption index, the more severe the problem

The higher the value of the governance indicator, the better the quality of that dimension

the challenges facing good governance and the potential causes of corruption. They can also help researchers to understand the link between governance, corruption, and quality of service delivery. In the present section, we offer two examples of the potential explanatory power that these indicators have.

Consider first the following research question: does the quality of a country's governance system affect access to public services? Using responses from 1123 Peruvian public officials and 1696 Peruvian households, Kaufmann et al. (2008) attempt to answer this question. First, they explore the costs of bad governance and the relative importance of various governance determinants in access to public services. The evidence based on the household data suggests that for certain basic services, low-income users pay a larger share of their income than wealthier ones. In addition, low-income users are more likely to be discouraged and to not seek a service than wealthier ones, especially when in need of a basic service (such as education, water, or the police). Thus, bribery appears to penalise poorer users twice, by acting both as a regressive tax and as a discriminating mechanism in access to basic services.

Kaufmann, Monteriol, and Recanatini then analyse household behaviour when attempting to obtain a public service. The survey data available allows empirical exploration to be carried out at two levels: at the household level and at the public agency level, as in Kaufmann et al. (2002). The analysis suggests that individual characteristics, such as education and age, matter in the decision of whether or not to seek a service when it is needed: higher educational attainment and middle-income users are associated with a higher probability of not seeking a public service when corruption is present. Governance, measured as trust in state institutions, also influences user behaviour. Finally, knowledge of reporting corruption and social networking increase the probability of being discouraged.

By exploiting the richness of the data and its disaggregated nature, the three authors complement the household-level analysis with a public official analysis. At the agency level, the public official data allows agency-level indices to be constructed for both access to public services and institutional factors, namely, the existence of audit mechanisms, the quality of rules governing each agency, individuals' understanding of the agency mission, availability of resources, and the extent of different forms of corruption, as reported by public officials. The Ordinary Least Square estimation results suggest that corruption reduces access to services, while voice mechanisms and understanding of the agency's mission increase it.

The previous example focused on the impact of the quality of governance on access to services within the same country. The indicators presented can also be employed to explore the potential causes of corruption and provide some empirical support for some policy tools to address the problem. In this case, the agency-level indicators based on responses from public officials can

help understand which features of the agency influence corruption and how to curb corruption inside each agency.

Recanatini (2016) uses the data collected through the surveys of the public employees working inside each public agency and of the agency's customers (households or firms) for ten countries. Such disaggregated analysis allows understanding of the institutional and governance factors that can influence corruption within a country.

The first step of the analysis in the paper is to explain each corruption assessment (by public officials, households, and firms) in terms of a large set of agency characteristics, which include the sector that is the main focus of the agency's operations (agriculture, finance, manufacturing, infrastructure, education, health, internal order or national security, etc.); the individuals with whom the agency interacts (households, firms, foreigners); and whether the agency makes large purchases from suppliers, is in charge of revenue collection, has a head that is popularly elected, is a municipality, is part of the national executive, or is a judicial agency.

The analysis finds several indications that corruption is greater where demand for the agency's services is relatively inelastic because there are few or no alternatives to the valuable service that the agency provides. This is most true of agencies that are in charge of internal order and national security, the judiciary, and those in charge of revenue collection. A finding that agencies dealing with domestic firms seem relatively more corrupt than those dealing with foreigners is also consistent with this interpretation. Similarly, public officials perceive utilities (which are mostly monopoly providers) as significantly more corrupt and firms report paying bribes to them relatively more frequently. Furthermore, all the respondents share the view that corruption is more prevalent in agencies whose head is popularly elected. Within this class of agencies, those that have legislative power (parliaments and local assemblies) and local governments are considered more corrupt than the central government. Finally, another strong result is that agencies dealing with infrastructure projects are significantly more corrupt.

In a second step, Recanatini (2016) focuses on a few specific aspects of the internal organisation of each agency. In particular, she focuses on the following three mechanisms: having decisions regularly audited by external or internal auditors ("audit"), basing personnel decisions on criteria regarding merit and professional competence ("merit"), and maintaining open and transparent procedures for budget and personnel decisions ("openness"). Studying how these internal features vary across agencies provides useful information on possible channels through which external forces or operational sectors influence the overall level of corruption in the agency. For example, the significantly better "audit" mechanisms that characterise public agencies dealing with the financial sector may explain why this sector is not seen as particularly corrupt.

The third and final step of the analysis is to address the more challenging question of whether internal organisation features have a causal effect on corruption. In other words, whether policy-makers can raise "audit", "merit", and "openness" over and above the level determined by the agency characteristics and expect corruption to fall. Recanatini's evidence suggests that the internal organisation measures used—audit, merit, and openness—do have an impact on the level of corruption reported within each public agency. Thus, through the use of agency-specific indicators of governance, Recanatini (2016) provides evidence that, on the demand side, corruption is more prevalent among agencies that provide services to firms (rather than households) and among those that provide an exclusive service for which there is no alternative. On the supply side, the internal organisation of the agency is a major determinant of corruption.

Conclusions

The present chapter has discussed the importance of moving beyond aggregate indicators of governance and corruption when focusing on policy design or exploring specific research questions related to measuring governance and corruption. The existing aggregate indicators fail to capture the heterogeneity of these phenomena and the way in which exchanges and transactions are carried out in practice despite the existing rules and regulations. The approach introduced in this chapter highlights the possibility of using survey data to overcome these limits and to construct indicators of governance and corruption that can help better understand these phenomena at the agency and at the regional levels. Moreover, survey data focused on experience (rather than perceptions) can help provide a more realistic assessment of the governance system of a country.

These indicators have mainly been used for policy purposes, to support the development of reform plans to improve governance and address corruption at the country level. However, the chapter has offered two examples of the potential use of the indicators for research purposes. The two illustrations discussed evidence the importance of specific organisational features in explaining the quality of governance and the extent of corruption in practice. The single most important direction for future research is to gain a better understanding of why internal organisational design varies so much across agencies. Why do some agencies adopt more open and transparent procedures or rely on external auditors while others do not? It would be important to be able to answer this question as it would take practitioners one step closer to making policy recommendations that are feasible and likely to be implemented in practice.

Annex 8.1

Table 8.4 Survey questions used to construct governance variables

Governance variable	Questions available in the survey
Access	Our agency's services are accessible by the poor
Service performance	Service delivery in your institution is of high quality
	Services in your institution are offered at relatively low cost
	Service delivery in your institution is fully satisfactory to the user
Audit	Decisions on personnel management are regularly audited by some internal unit
	Decisions on personnel management are subject to external audits
	Decisions on budget management are regularly audited by some internal unit
	DecDecisions on budget management are subject to external audits
	Decisions on service delivery/contracts are regularly audited by some internal unit
	Decisions on service delivery/contracts are subject to external audits
	Transactions are supported by hard-copy receipts to help auditing
Enforcement of rules	The policy/guidelines/regulations on the administration of personnel management are strictly enforced
	The policy/guidelines/regulations on the administration of budget management are strictly enforced
	The policy/guidelines/regulations on the administration of service management/public contract management are strictly enforced
Supervision of rules	The policy/guidelines/regulations on the administration of personnel management are well supervised
	The policy/guidelines/regulations on the administration of budget management are well supervised
	The policy/guidelines/regulations on the administration of service management/public contract management are well supervised
Meritocracy	Percentage of cases where decisions on personnel management issues are based on professional experience/merit/ performance
	Percentage of cases where decisions on personnel management issues are based on level of education
Mission	All levels of public servants have a clear understanding of the agency's objectives and strategies
	All levels of public servants identify with and are involved with the agency's objectives and strategies
	I clearly understand my institutional roles and responsibilities and which duties are comprised in each role

(*continued*)

Table 8.4 (continued)

Governance variable	Questions available in the survey
Openness	Decisions on personnel management are announced and opened to the inside of the institution (and also to the outside if applicable)
	Decisions on budget management are announced and opened to the inside of the institution (and also to the outside if applicable)
	Decisions on service delivery/performance of daily tasks/public contracts are announced and opened to the inside of the institution (and also to the outside if applicable)
	The agency's financial status is regularly disclosed to the public
Politicisation	Decisions on personnel management are based on political connections/party affiliations/political pressure
	Decisions on budget management are based on political connections/party affiliations/political pressure
	Decisions on service management (or bidding processes) are based on political connections/party affiliations/political pressure
Quality of rules	The policy/guidelines/regulations on the administration of personnel management are written and formal
	The policy/guidelines/regulations on the administration of personnel management are simple, clear, and easy to understand
	The policy/guidelines/regulations on the administration of personnel management do not add too many administrative steps
	The policy/guidelines/regulations on the administration of budget management are written and formal
	The policy/guidelines/regulations on the administration of budget management are simple, clear, and easy to understand
	The policy/guidelines/regulations on the administration of budget management do not add too many administrative steps
	The policy/guidelines/regulations on service management/ bidding processes are written and formal
	The policy/guidelines/regulations on service management/ bidding processes are simple, clear, and easy to understand
Resources	The quantity (and quality, if applicable) of the agency's resources is adequate
	The agency has adequate personnel and they are adequately trained
	The agency has adequate office supplies/computers
	The agency has adequate space/offices

(*continued*)

Table 8.4 (continued)

Governance variable	Questions available in the survey
Transparency	Decisions on personnel management are made transparently (everybody knows who was designated, promoted, etc.)
	Decisions on budget management are made transparently (everybody knows what was spent, etc.)
	Decisions on service delivery/performance of daily tasks/public contracts are made transparently (everybody knows what was decided)
Voice	We all consider that citizens and users are our clients
	Clearly defined mechanisms exist to ask users about their needs
	Clearly defined mechanisms exist so that users can express their preferences, suggestions, and complaints
Rotation public/ private sectors	The tendency of the public officials in your institution (director, executives, professionals, technicians, assistants) to change jobs and rotate from the public to the private sector
	The tendency of persons to rotate from the private sector to the public sector
Wage satisfaction	The percentage of employees who are very satisfied or somewhat satisfied with their wages and benefits

Source: Own compilation

Table 8.5 Governance and corruption indicators by the province, Zambia

Indicators	Province									Whole Country
	Central	Copper belt	Eastern	Luapula	Lusaka	Northern	North-western	Southern	Western	
Overall corruption	24	27	40	37	25	36	31	21	16	28
Administrative corruption	25	31	45	43	30	48	27	33	20	33
State capture	18	21	36	34	21	32	25	16	12	32
Corruption in budget	36	44	43	40	33	43	46	30	15	37
Corruption in public contracts	37	34	47	47	35	49	35	24	22	36
Corruption in personnel (1)	14	24	37	26	20	23	18	18	9	22
Corruption in personnel (2)	41	41	54	50	38	53	45	25	34	41
Accessibility for poor	69	63	68	69	67	67	61	71	70	67
Audit mechanisms	64	63	66	55	68	60	55	58	68	63
Enforcement of rules	56	49	60	53	57	57	49	56	58	55
Politicization	38	42	40	39	38	36	38	39	36	39
Quality of rules	60	54	56	51	62	61	52	58	61	57
Resources	56	40	59	40	50	51	38	48	53	48
Transparency	54	51	50	39	58	57	34	55	54	52
Citizen voice	62	58	57	53	55	57	56	59	55	57
Meritocracy	64	62	68	56	65	65	57	68	75	64

Source: Own compilation

The indicators above take values between 0 and 100. To interpret them it should be kept in mind that:

The higher the value of the corruption index, the more severe the problem

The higher the value of the governance indicator, the better the quality of that dimension

Table 8.6 Governance and corruption indicators by institution, Sierra Leone

	Audit	Enforcement of Rules	Resources	Transparency	Citizen Voice	Wage Satisfaction
Presidential affairs	52	74	51	55	59	17
Ministry of Finance	57	68	52	54	60	25
Quasi-independent institutions	69	85	41	63	78	32
Ministry of Trade and Industry	64	84	60	59	63	19
Ministry of Internal Affairs	60	76	57	61	76	0
Ministry of Energy and Power and NPA	72	92	62	61	69	25
Ministry of Foreign Affairs and International Cooperation	50		74	54	50	30
Ministry of Development and Economic Planning	54	72	69	57	58	25
Ministry of Social Welfare, Gender & Children's Affairs	59	71	53	56	58	13
Ministry of Health and Sanitation	54	71	53	55	65	21
Ministry of Information and Broadcasting	55	68	51	51	60	18
Ministry of Education, Science and Technology	54	69	51	55	67	26
Ministry of Youth and Sports	56	76	50	52	72	25
Ministry of Labor and Industrial Relations	61	71	48	56	64	25
Ministry of Agriculture and Food Security	60	69	46	58	66	16
Ministry of Local Government and Municipal and District Councils	58	72	54	57	63	15
Ministry of Transport and Communications	64	77	46	57	66	31
Ministry of Works and Public Maintenance	53	67	52	53	68	5
Ministry of Lands, Housing and Country Planning			54	57	56	13
Ministry of Mineral Resources	60	67	68	55	55	21
House of Parliament	54	53	41	59	50	13
Office of the Attorney General, Minister of Justice and Director of Public Prosecutor	60	75	49	55	67	25
Traffic Police and Road Transport Authority	41	74	64	54	63	34
Supreme Courts	64	78	55	56	72	43
Local Courts	61	68	41	53	77	36
SALWACO/GVWC and SIERRATEL	56	73	53	57	58	35
SALPOST	61	66	46	54	60	28
Port Authority	53	66	56	54	59	42
SL National Tourist Board and SL Standards Bureau	44		47	49	51	63
Police and Prisons	59	68	57	53	66	25
Post-conflict institutions	60	71	47	54	64	44
WHOLE COUNTRY	58	71	52	55	66	25

Source: Own compilation
Colour coding to facilitate reading of the results:
Agencies highlighted in *red* score well below the national indicator
Agencies highlighted in *green* score well above the national indicator

Notes

1. For the purposes of this chapter, governance is defined as "the traditions and institutions by which authority in a country is exercised" (Kaufmann et al. 1999a, b, p. 1) and corruption is defined as the use of the power of public office for personal gain. Corruption is taken to include all (and only) activities in which "public officials, bureaucrats, legislators and politicians use powers delegated to them by the public to further their own economic interests at the expense of the public good" (Jain 2001, p. 73). This definition embraces many different forms of corruption—from administrative corruption to "state capture"—when powerful groups buy influence and shape laws to their benefit.
2. This is in line with the "resistance to reforms" argument first highlighted by Fernandez and Rodrik (1991).
3. See Kaufmann et al. (1998) for a description of the original approach and its implementation in Latvia, Georgia, and Albania.

4. Recanatini (2012) offers a detailed discussion of these tools and their applicability for policy reform.
5. To test the robustness of our approach, we also calculate agency indices of governance as mean values of the individual responses.

References

Bardhan, P. (1997). Corruption and Development. *Journal of Economic Literature, 35*(3), 1320–1346.

Broadman, H., & Recanatini, F. (2002). Seeds of Corruption. *MOCT-MOST: Economic Policy in Transitional Economies, 11*, 359–392.

Charron, N. (2013). *From Aland to Ankara: European Quality of Government Index* (Working Paper Series 2013:11). Gothenburg: Quality of Government Institute, University of Gothenburg.

Fisman, R., & Gatti, R. (2002). Decentralization and Corruption: Evidence Across Countries. *Journal of Public Economics, 83*, 325–345.

Fernandez, R., & Rodrik, D. (1991). Resistance to Reform: Status Quo Bias in Presence of Individual Specific Uncertainty. *American Economic Review, 81*(5), 1146–1155.

Friedman, E., Johnson, S., Kaufmann, D., & Zoido-Lobaton, P. (2000). Dodging the Grabbing Hand: The Determinants of Unofficial Activity in 69 Countries. *Journal of Public Economics, 76*, 459–493.

Gupta, S., Davoodi, H., & Alonso-Terme R. (1998). *Does Corruption Affect Income Inequality and Poverty?* (International Monetary Fund Working Paper, WP/98/76). Washington, DC: International Monetary Fund.

International Monetary Fund. (1996). *Summary Proceedings of the Fifty-First Annual Meeting of the Board of Governors October 1–3, 1996*. Washington, DC: International Monetary Fund.

Hunt, J., & Laszlo, S. (2005, September). *Bribery: Who Pays? Who Refuses? What Are the Payoffs?* (National Bureau of Economic Research Working Paper Series, 11635). Cambridge, MA: National Bureau of Economic Research.

Jain, A. (2001). Corruption: A Review. *Journal of Economic Surveys, 15*(1), 71–121.

Johnson, S., Kaufmann, D., McMillan, J., & Woodruff, C. (2000). Why Do Firms Hide? Bribes and Unofficial Activity After Communism. *Journal of Public Economics, 76*, 495–520.

Johnston, M., & Kpundeh, S. (2004, December). *Building a Clean Machine: Anti-Corruption Coalitions and Sustainable Reform* (World Bank Policy Research Working Paper, 3466). Washington, DC: World Bank.

Kaufmann, D. (2000). Governance and Anti-Corruption. In V. Thomas (Ed.), *The Quality of Growth* (pp. 135–168). New York: Oxford University Press.

Kaufmann, D., Pradhan, S., & Ryterman, R. (1998). *New Frontiers in Diagnosing and Combating Corruption* (PREM Note #7). Washington, DC: World Bank.

Kaufmann, D., Kraay, A., & Zoito-Lobaton, P. (1999a). *Governance Matters* (World Bank PRD Working Paper No. 2196). Washington, DC: World Bank.

Kaufmann, D., Kraay, A., & Zoito-Lobaton, P. (1999b). *Aggregating Governance Indicators* (World Bank Policy Research Working Paper No. 2195). Washington, DC: World Bank.

Kaufmann, D., Mehrez, G., & Gurgur, T. (2002). *Voice or Public Sector Management? An Empirical Investigation of Determinants of Public Sector Performance Based on a Survey of Public Official* (World Bank Research Working Paper). Washington, DC: The World Bank.

Kaufmann, D., Montoriol, J., & Recanatini, F. (2008). *How Does Bribery Affect Public Service Delivery? Micro-Evidence from Service Users and Public Pfficials in Peru* (World Bank Policy Research Working Paper, No. 4492). Washington, DC: The World Bank.

Kisunko, G., Knack, S., & Islam. (2013). Russian Federation: National and Regional Trends in Regulatory Burden and Corruption. World Bank Policy Note, February 2013. Washington, DC: World Bank.

Knack, S., & Anderson, J. (1999). *Is "Good Governance" Progressive? Institutions, Inequality, and Poverty Reduction*. Paper presented at the 1999 Annual Meeting of the American Political Science Association, Atlanta, Georgia.

Kpundeh, S., & Stapenhurst, R. (1999). *Curbing Corruption: Toward a Model for Building National Integrity*. Washington, DC: Economic Development Institute, The World Bank.

Mauro, P. (1995). Corruption and Growth. *Quarterly Journal of Economics, 110*(3), 681–712.

Recanatini, F. (2012). Assessing Corruption at the Country Level. In A. Graycar & R. G. Smith (Eds.), *Handbook of Global Research and Practice in Corruption* (pp. 34–64). Cheltenham: Edward Elgar.

Recanatini, F. (2016). *Why Are Some Public Agencies Less Corrupt than Others? Lessons for Institutional Reform from Survey Data*. Background paper for the World Development Report 2016. Washington, DC: World Bank.

Svensson, J. (2003). Who Must Pay Bribes and How Much? *Quarterly Journal of Economics, 118*(1), 207–230.

Svensson, J. (2005). Eight Questions About Corruption. *Journal of Economic Perspectives, 19*(3), 19–42.

Tanzi, V., & Davoodi, H. (1997). *Corruption, Public Investment and Growth* (IMF Working Papers, WP/97/139). Washington, DC: International Monetary Fund.

Treisman, D. (2000). The Causes of Corruption: A Cross-National Study. *Journal of Public Economics, 76*(3), 399–457.

World Bank. (2001). *Voices of the Misgoverned and Misruled: An Empirical Diagnostic Study on Governance, Rule of Law and Corruption for Peru*. Washington, DC: The World Bank.

World Bank. (2003). Governance and Anticorruption Diagnostic Report for Sierra Leone, produced in collaboration with the Government of Sierra Leone, 2003. Washington, DC: World Bank.

World Bank. (2004). *Making Services Work for the Poor*. World Development Report. Washington DC: World Bank.

9

Government at a Glance: A Dashboard Approach to Indicators

Guillaume Lafortune, Santiago Gonzalez, and Zsuzsanna Lonti

Introduction

Since the mid-1980s, there has been increasing interest in measuring the performance of governments, both at the national and international level. There have been several reasons for this rise in performance measurement in the public sector. Declining economic growth—starting with the impact of the 1970s oil crisis—rising unemployment, and an aging population put additional pressures on public finances in most OECD countries. Scarce public resources call for more and better information on how efficiently and effectively these resources are used. Increasing globalisation has also played an important role by putting new emphasis on government policies, practices, and performance, creating increased competition among governments to attract foreign investment. In the international aid community, there has also been an increased demand to base aid allocation on better information on the efficiency and effectiveness with which recipient governments are able to utilise the resources they receive. International governance indicators and

This chapter is based on the OECD report "OECD (2015), Government at Glance 2015, OECD Publishing, Paris 10.1787/gov_glance-2015-en" and "OECD (2015), OECD Regulatory Policy Outlook 2015, *OECD* Publishing, Paris. DOI: http://dx.doi.org/10.1787/9789264238770-en". Any additional opinions expressed or arguments employed herein are solely those of the authors and do not necessarily reflect the official view of the OECD or its member countries.

G. Lafortune (✉) • S. Gonzalez • Z. Lonti
OECD, Paris, France

D.V. Malito et al. (eds.), *The Palgrave Handbook of Indicators in Global Governance*,
https://doi.org/10.1007/978-3-319-62707-6_9

scoreboards have therefore been increasingly used by investors, aid donors, and structural fund providers to evaluate the stability and transparency of governments and the ease of conducting affairs in different countries (Arndt and Oman 2006).

This period has also given rise to New Public Management (NPM), with its insistence on the need for better management practices in the public sector by importing private sector management practices—among them performance measurement and management (Osborne and Gaebler 1992). Increasing globalisation also directly applies to public management practices, with countries learning from each other and copying each other's practices (Hood 1991). At the same time, rising citizen expectations of more efficient and better quality public services also call for better reporting on public sector performance.

Performance information on how governments work can serve a number of stakeholders: first, the users of public services (e.g. by giving parents information on how their children's school performs compared to other schools); second, the providers of the services (e.g. by helping the school improve its performance); third, funders (e.g. by helping the Ministry of Education know how well the resources it spends are utilised by individual schools); and, last but not least, citizens (e.g. by informing them about how efficiently and effectively public money is spent). As a result, performance measurement can provide information to improve the performance of an individual unit by allowing it to benchmark itself against the performance of other similar units or by comparing its performance over time. It can also provide vital information to governments on how their resources are spent and can improve their accountability to the users of their services and to citizens in general for its actions and performance.

In this chapter, the focus is on internationally comparative measures of government performance and on one of these in particular, the OECD's public governance indicator set, which is included in its publication entitled Government at a Glance (G@G). The objective of the chapter is to contribute to the existing literature and debates on the utility of governance indicators and demonstrate the usefulness of the dashboard approach used in G@G to accurately inform public decision-making.

The chapter is divided into three sections. The first section introduces the dashboard approach that is followed by G@G, explaining its origins, its major objectives, the type of indicators included, and the role of OECD member countries in the process. The second section introduces the framework—five key criteria—which is used to evaluate the quality of the indicators published in G@G. These five criteria are designed to help respond to the objective of the indicator set to provide relevant and robust evidence to member countries

to inform public sector reforms and decision-making. The third section applies the evaluation framework to various indicators included in G@G to demonstrate the strengths and sometimes some weaknesses of the G@G indicator set.

What Is Government at a Glance?

G@G is a public governance indicator set published biannually since 2009 by the OECD and contains information on its member and partner countries focusing on how governments work and their performance from an internationally comparative perspective. Its major objectives are to allow countries to benchmark their performance against other countries, to allow them to measure their own progress over time, to provide evidence to policy-makers, and to allow countries to identify where further progress is needed or if there are problem areas in order to address these areas.

G@G follows a dashboard approach. This means that it includes narrowly defined and often detailed indicators on government inputs, activities, outputs, and outcomes without creating any super-aggregations of this set of indicators (e.g. super-composites) aimed at summarising the performance of activities (e.g. digital government performance, regulatory governance) or government as a whole into a single figure.

G@G is not constructed on the basis of an explicit definition of what good governance is, but good governance principles are implicitly reflected in the choices of the indicators presented and how these indicators are built. The various areas of public management—from budgeting and human resource management (HRM) practices to open government and integrity—reflect what matters in good governance. Some of the key indicators—such as those on regulatory quality management—reflect common understanding among the OECD member countries of best practices in a particular governance area. For example, the regulatory quality management indicators reflect the 2012 OECD Recommendation on Regulatory Policy and Governance.

The major target audience of G@G is politicians, public policy-makers, and practitioners, but it also aims to reach a wider audience, including opinion leaders, academics, students, public interest groups, and, not least, interested citizens. The country coverage of the publication is growing from edition to edition, for example, in 2009, it covered 30 OECD member countries but in 2015 it included 34 member countries plus two countries in the OECD accession process (Colombia and Latvia) as well as partner countries (China, Brazil, India, Indonesia, and South Africa) and Russia, reaching over 43 countries.

The Role of OECD Networks in G@G

G@G was created on the initiative of the OECD member countries through its Public Governance Committee (PGC), which identified the need for an internationally comparable set of public governance indicators that would allow it to benchmark progress on public management reforms. It also builds on the work programme and expertise of the various networks operating under or beside the PGC, such as the Senior Budget Officials Network, the Regulatory Policy Committee, and the Public Employment and Management Network, to name a few. These networks and their members are instrumental in the decisions on the kinds of indicator to be included in G@G and the quality processes followed.

The members of the networks are government officials who are experts in their fields. They are involved in all the key steps of indicator development (topic selection, data collection method, survey design, data cleaning, and validation of the results). They are also in charge of responding to surveys and identifying the right respondents. In this way, the indicators included in G@G reflect the official position of governments. However, the information provided by these national experts is reviewed by topic experts at the Secretariat and any discrepancies are brought to the attention of the relevant national government. The results are also discussed by the networks and the PGC.

Government officials are the respondents for most G@G indicators and they provide either official government statistics (e.g. on public finances and employment), government administrative data (e.g. on remuneration for ICT), or the country's own assessments of its practices and procedures. While the surveys include standardised definitions, bias can occur as countries may interpret and answer the questions differently and/or they may not answer the questions completely objectively. In order to try and minimise bias, the surveys contain extensive guidance on how to interpret the questions and fill out the responses.

The Indicators Included in G@G

G@G includes indicators on the whole "production chain" of government activities, from inputs to public management and governance processes to outputs and outcomes. A combination of these indicators also allows efficiency and effectiveness measures in some areas to be calculated. Putting this information in its political and cultural context is also important, so G@G also publishes information on key institutional features of governments.

There are indicators of the size of government in terms of revenues, expenditures, employment, and remuneration (or compensation), as well as on key

public management and governance practices, such as HRM, budgeting, regulatory management, digital government, open government, public procurement, and integrity. These latter indicators provide insights into the government's administrative capacity to deal with current and future policy challenges. As for information on the performance of governments, it is understood in a broad sense, encompassing indicators both on the performance of the various governance processes—such as budgeting practices and integrity tools—and the performance of key policy sectors, such as education, health, justice, and tax administration. These indicators include both efficiency and effectiveness, for health and education, and access to and quality of services, for health, education, and justice. In addition, G@G also includes indicators on core government results, such as trust in government, fiscal sustainability, and the role of government in reducing inequalities (Fig. 9.1).

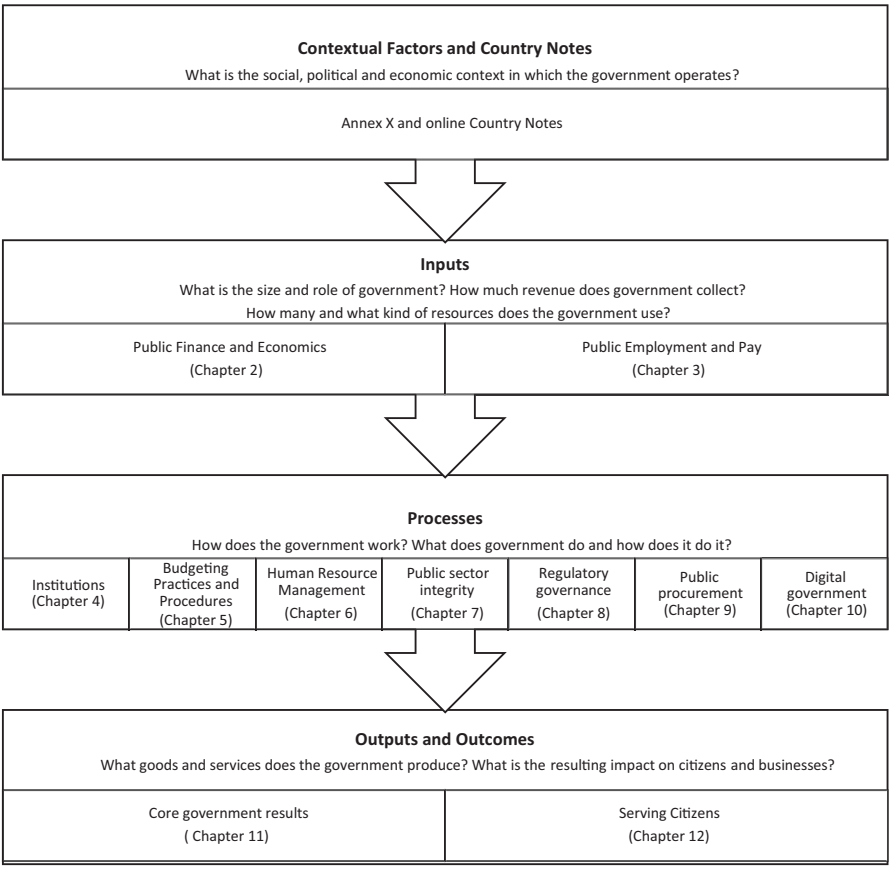

Fig. 9.1 Conceptual framework of government at a glance 2015. Source: OECD Government at a Glance (2015a)

The number of indicators in G@G changes from edition to edition. This is a reflection of several factors working together. First, economic developments—such as the crisis and its aftermath—have brought new topics to the forefront and these are reflected in indicators, such as fiscal sustainability, government investment, and the need for better strategic foresight, to name just a few. Second, indicators on government performance in the form of core government results and on policy sector outputs and outcomes—for example, education, health, and justice—have been included since the second edition. This reflects the results of a debate regarding the scope of the indicator set—whether it should only focus on the narrowly defined public administration or the whole public sector. While the 2009 edition reflected the first view, in response to requests from the OECD member countries, the focus since 2011 has been more on the whole public sector (OECD 2009, 2011, 2013a, 2015a). Third, there is institutional learning at play: as country differences are described, the need to include more indicators that can explain them arises. As a solution, there are three types of indicators in G@G: (a) core, (b) periodic, and (c) special features.

Core indicators are part of every release and include government revenues, expenditures, employment, public governance practices related to compliance with core public values such as integrity and transparency, core government results, and major institutional features. Since 2015, indicators on access to public services and on their quality—for the education, health, and justice sectors—have also been part of the core indicator set.

Periodic indicators are primarily "process" indicators, dealing with public management and public governance practices which are less prone to change, such as budgeting practices, HRM practices, and regulatory quality management practices. However, they also include data on remuneration levels and employment structures in key central government occupations. They are measured periodically by means of OECD surveys and are included in every second or third edition of the publication.

Finally, there are indicators that are included as special features, addressing topical issues at the time of publication. For example, the 2011 edition featured new data on green procurement, while in 2015 indicators on the impact of the fiscal crisis on HRM practices were included.

Data Sources Used in G@G

G@G uses many data sources, although most of the data are collected by the OECD. Data are also sourced from the International Labour Organisation (ILO) on public employment, Eurostat on e-government, the International

Monetary Fund (IMF) on public finance for non-member countries, the European Commission for the Efficiency of Justice (CEPEJ), the World Justice Project (WJP) also on the justice sector, and the Gallup World Poll on trust in government and public institutions.

The data sources used in G@G can be divided into three broad types:

(1) Population-based (household) surveys;
(2) Expert surveys (with the experts possibly including government officials and non-government stakeholders);
(3) Administrative data.

Each of these broad types of data source has strengths and weaknesses. They can be described as being more or less subjective, with general population-based opinion surveys considered the most "subjective", administrative data being the most "objective", and expert surveys falling somewhere in between.

Household Surveys

Large household-based opinion surveys carried out by some organisations (Gallup World Poll, World Value Survey, and various barometers) collect information on the perceptions of populations regarding their level of satisfaction with and confidence in their national governments and their services. They provide useful information centred on citizens. In most cases, time series are available to compare trends over time. These population-based opinion surveys may sometimes be the only source of information to assess the level of public confidence and trust in the government in general or in specific government services.

However, these international perception-based surveys must be interpreted with caution for a number of reasons. Because they are costly, the sample sizes are often fairly small (1000 people per country for the Gallup World Poll), which means that the results have high margins of error (sampling errors). The confidence intervals are not always shown and so are not included in G@G. National surveys often have a larger sample size, but for the purpose of cross-country comparisons, they need to be harmonised ex ante in terms of the survey items (questions and answers) and methodology in order to provide comparable results. There can also be cultural biases in how people in different countries respond to survey questions, reflecting national "traits". Subjective data rely on respondents' understanding of the questions and as a result they can be strongly dependent on semantics. Finally, people's opinions

can change somewhat frequently in relation to changes in the context or in public opinion without any relation to the structural phenomenon that the indicator aims to measure (e.g. as a result of media campaigns raising public awareness of particular issues).

Expert Surveys

In G@G, most of the data are based on responses from the government officials who are responsible for data collection on the specific area/topic in their national administration (chief information officers, chiefs of staff, senior officials in cabinet offices, finance ministries, etc.).

The responses collected in the surveys sent to designated government officials are generally used to represent the official position of the government. These high-level officials typically coordinate internally and consult with other relevant units and sections within their administrations which can assist them in providing the responses to the surveys. As G@G is a statistical publication comparing the practices of public administrations, this type of information is useful to assess the evolution of practices and the design, implementation, and impacts of specific reforms in OECD countries. These surveys also often ask government officials to disclose administrative data related to the subject matter.

However, there are also some limitations to using the responses provided by government officials. Generally, the number of experts who provide responses to the surveys is rather limited. These contact points do not always have the ability to coordinate effectively inside their administration, resulting in missing answers or unreliable responses (which are generally corrected during the data-cleaning process). In addition, while these surveys are useful to collect information on the evolution of practices, they are typically less useful to collect information on the actual results and impacts of public policies on certain populations.

Administrative Data

In G@G, administrative data[1] are used, for instance, to measure government inputs (financial and human) and the efficiency of judicial systems (e.g. the length of judicial processes in days), the efficiency and technical efficiency of healthcare systems (average length of stay in hospitals, share of cataract surgeries carried out as day cases), and the quality of healthcare systems (cervical cancer 5-year survival rate). These efficiency and quality indicators are based

on a harmonised methodology and definitions collected from court and hospital records, which are then aggregated and collected by the CEPEJ and the OECD healthcare surveys (in these two cases).

While in theory high-quality administrative data provide more reliable measures than subjective data (which are inherently less replicable due to the respondents' interpretations of the questions), they are also subject to a number of limitations. First, there are limitations in data reporting and validity (e.g. "gaming" to over-report positive activities to obtain rewards or to under-report negative activities to avoid sanctions). Second, the use of administrative data requires significant efforts to harmonise definitions of concepts and their scope ex ante. Thus, there is a high cost (in both money and time) involved in defining and conceptualising the key measures of performance beforehand in collaboration with the survey respondents and also in ensuring that methodologies are harmonised at the national level (in this case, at the court and hospital levels) to ensure the comparability of results across countries.

Individual Indicators Versus Composite Indicators

Most of the indicators in G@G are individual indicators.[2] However, some concepts of government activities are multidimensional and cannot be accurately measured/captured by single quantitative or qualitative indicators. The extent to which HRM practices are career-based or position-based, the extent of the use of performance indicators (or medium-term budgeting frameworks) in central/federal budgeting, or the extent to which public officials disclose their private interests and make the information publicly available are multidimensional concepts which can only be measured by aggregating single measures into larger composites based on solid theoretical grounds.

Composite indicators correspond to aggregations/combinations of individual indicators to measure one larger concept. In G@G, composite indicators have mainly been developed in the areas of HRM, regulatory practices and procedures, and budgeting practices and procedures. For instance, a composite indicator measuring the extent to which governments use performance information in the budgeting process was developed using several responses from the OECD Budgeting Practices and Procedures survey.

The OECD collaborates with working parties and expert groups to ensure that the choice of underlying variables and weightings is strongly anchored in theory and reflects a common understanding of the concept measured. For instance, the composite indicators on regulatory practices in the 2015 edition

of G@G (iReg) were developed after years of close collaboration with the OECD Regulatory Policy Committee. These composites were designed to measure the level of implementation of the 2012 OECD Recommendations on Regulatory Policies, such as stakeholder engagement in shaping primary and subordinate regulations at the central/federal levels of government.

The statistical soundness of composite indicators is ensured by closely following the various steps identified in the Handbook on Composite Indicators developed jointly by the OECD and the European Commission (Nardo et al. 2005). Three "key" statistical tests are systematically carried out and provided in the annexes of the G@G publication (which is also available online):

(1) Confirmatory principal component factor analysis;
(2) Cronbach alpha testing (to measure the degree of correlation among the variables underlying each factor);
(3) Sensitivity analysis (Monte Carlo simulation).

While composite indicators can be useful tools for policy-making and public sector reforms, they can also be misleading tools when not constructed rigorously (Table 9.1). Since they involve assumptions, they need to be carefully assessed to ensure their relevance and robustness. By showing the "big picture", they may invite users (including policy-makers) to draw simplistic analytical or policy conclusions.

A Framework to Evaluate the Quality of Public Governance Indicators

This second section provides a set of criteria to evaluate the relevance and robustness of public governance indicators. The criteria below have been developed over time to assess the quality of the indicators that are included in G@G. There are five in total: three to evaluate the degree of **relevance** of governance indicators and two to evaluate their **robustness** (or statistical soundness) (see Table 9.2 for a summary of this section).

Evaluating the Relevance of Governance Indicators

Relevance corresponds to the degree to which indicators serve a clear purpose and provide useful information that can be used to guide public sector reforms in member countries. The target audience of the publication is primarily

Table 9.1 Strengths and weaknesses of composite indicators

Strengths	Weaknesses
Can summarise complex, multidimensional realities with a view to supporting decision-makers	May send misleading policy messages if poorly constructed or misinterpreted
Are easier to interpret than a battery of many separate indicators	May invite simplistic policy conclusions
Can assess progress of countries over time	May be misused, for example, to support a desired policy, if the construction process is not transparent and/or lacks sound statistical or conceptual principles
Reduce the visible size of a set of indicators without dropping the underlying information base	The selection of indicators and weights could be the subject of political dispute
Make it possible to include more information within the existing size limit	May disguise serious failings in some dimensions and increase the difficulty of identifying proper remedial action, if the construction process is not transparent
Place issues of country performance and progress at the centre of the policy arena	May lead to inappropriate policies if dimensions of performance that are difficult to measure are ignored
Facilitate communication with general public (i.e. citizens, media, etc.) and promote accountability	The linear aggregation method used in most composite indicators assumes that the components are compensatory
Help to construct/underpin narratives for lay and literate audiences	
Enable users to compare complex dimensions effectively	

Source: OECD and European Commission, *Handbook on Constructing Composite Indicators* (2009); Adapted from Saisana & Tarantola, 2002

decision-makers. Therefore, to be useful and relevant, it is fundamental that the indicator sets provided are:

(1) Action worthy;
(2) Actionable;
(3) Behavioural.

The first criterion to assess the relevance of an indicator is "action worthiness". An indicator should measure something that is important, in other words, something that is worth governments taking action and which is meaningful for policy-makers and the broader society (Kaufmann and Kraay 2007, p. 6; Arndt 2008). As far as possible, the indicator should be connected to a clear and valuable outcome (such as greater access to public services, greater societal equity, or greater efficiency).

Table 9.2 Criteria to assess the relevance and robustness of governance indicators at the international level

Criteria	Key questions
Relevance	
Important (action worthiness)	Are policy-makers and the broader society concerned about what is being measured?
Actionable	Can policy-makers meaningfully address the issue or problem by taking concrete actions?
Behavioural	Does it measure observable practices/experiences (de facto) or the existence of institutional settings/forms (de jure)?
Robustness (scientific soundness)	
Valid	Does it measure what it is intended to measure?
Reliable	Does it provide stable results across various populations and circumstances?

Source: Authors' own construction

The second criterion to evaluate relevance is "actionability". On the basis of the indicator, governments and policy-makers should know what actions they need to take in order to improve their performance. Indicators should provide useful and informative insights into the types of reform countries should engage in (Arndt 2008, p. 279). Actionability can be reduced when an indicator is highly sensitive to external factors, that is, if most of the observed changes in the indicator are explained by factors beyond the control of governments, it is not likely to be useful to monitor a government's progress in implementing public sector reforms.

The third criterion to assess relevance is whether the indicator is "behavioural". While measuring the existence of directives, laws and other institutional documents or bodies (de jure) provide some information on the legal frameworks in place, what really matters is whether they are actually implemented and what their impact is. In order to effectively inform public sector reforms, indicators should generally measure actual and observable facts, practices, and implementation (de facto). For instance, the existence of a specific anti-corruption law or corruption-prosecution agency does not necessarily imply lower levels of corruption in practice (Arndt and Oman 2006). Similarly, the imposition of fiscal rules driven by recent fiscal consolidation experiences has at best had a low marginal impact on the actual behaviours of countries, which is what really matters (Holt and Manning 2014, p. 722). Triangulation between de jure and de facto information, however, remains useful to identify potential gaps in implementation (OECD 2014).

Evaluating the Robustness of Governance Indicators

Robustness corresponds to the statistical soundness of indicators. The OECD has developed a framework to guide statistical activities in the organisation (OECD 2012a). The literature in social science on sound qualitative and quantitative research methods is also very large. In this chapter, we limit the discussion to two core measures of statistical soundness:

(1) Validity;
(2) Reliability.

A valid indicator measures precisely the concept it is intended to measure. For instance, a test that aims to measure the capacity of students to perform arithmetical operations but that contains only addition problems but no subtraction, multiplication, or division problems will not produce a valid measure (Carmines and Zeller 1979). The concept of statistical validity can be tested in various ways: (a) face validity looks at whether or not respondents and/or data users judge that the items are appropriate given the assessment's objectives; (b) convergence validity analyses the extent to which the measure correlates well with other proxy measures of the same underlying concept; and (c) construct validity analyses whether the measure performs in the way theory would suggest with respect to the construct being measured.

The second criterion to assess robustness is the degree of "reliability" of the indicator. To be reliable, the measure should produce consistent results when repeated across populations and settings and even when assessed by different people or at different times. Variability should result from changes in the subject of measurement rather than from artefacts of measurement (e.g. a change in the definition of the measure or, for rare events, a restricted sample size or a small number of cases). To assess the reliability of perception-based questions, several statistical tests can be carried out, such as test/re-test or split samples.

Applying the Evaluation Framework to G@G

This section applies the evaluation framework to the various sections of G@G. We use a 5-point scale—very low, low, medium/varying, high, and very high—to evaluate the relevance and robustness of the indicators included

in the various G@G editions. This is a rather indicative/experimental approach based on the authors' own views, the objective being to identify some of the strengths of the publication and areas where the quality of the indicators could be improved in future years.

Government Inputs (Human and Financial Resources)

The first type of indicator included in G@G is composed of the resources (human and financial) required by governments to produce goods and services. Most of the indicators included in these chapters are quantitative. In the case of public employment, the data are not collected directly by the OECD but are drawn from the ILO. The OECD has worked in close cooperation with the ILO to redesign the joint public employment questionnaires according to a harmonised methodology and definitions. Depending on the country's institutional structure, the respondents could be National Statistical Offices (NSOs), National Employment Agencies, or the relevant line ministry. The answers are national statistics, mainly based on administrative data, and consolidated by country respondents.

A unique feature of G@G when compared to other sets of governance indicators is the availability of information on the remuneration of civil servants. On two occasions, the OECD has collected data on the salary costs in key central government occupations in six core ministries (education, health, interior, finance, environment, and justice). The perspective adopted is broader than just asking about gross salaries as it also includes employers' social security contributions (including pension liabilities—funded and unfunded) and a correction for the time actually worked. The respondent in the OECD Survey on Compensation of Employees in Central/Federal Governments is an employee of a central government agency responsible for HRM, but in the case that there is no central repository of this information, the respondent has to collect it from the six ministries. This is an example of an OECD survey using access to government to collect administrative data from member countries.

Action Worthiness

The resources available shape the capacity of governments to respond to citizens' demands and adapt to emerging circumstances. Considering that the largest share of government resources are collected from citizens and enterprises through taxes, the general public has an interest in how public money is spent and especially in keeping the government accountable for its use of

public resources. Additionally, it is expected that governments should provide value for money and adhere to efficiency considerations.

Turning to public employment, across OECD member countries, public employment represents an average of 21.3% of total employment. In other words, one in every five employees in the economy works for the government. In principle, governments are interested in attracting the most qualified people with the most appropriate sets of skills to deliver high-quality goods and services to their citizens. The public employment indicators in G@G provide an overview of the size of government employment. The data on public sector employment—which include employment at all levels of government, social security funds, and public corporations, separated for women and men—are available on a biannual basis. However, more detailed internationally comparable data do not exist on the policy sectors in which public employees work or on their characteristics, for example, age, education level, and occupational distribution. While society has an interest in how many people work in the public sector, it also expects the public sector workforce to reflect the diversity in the broader society. An inclusive public sector cannot only guarantee that all groups within society are represented, but can also better tailor the provision of public services to different groups. The action worthiness of the public employment indicators could be improved by adding information on diversity. Future editions of G@G will incorporate this type of information.

A first step in this direction has been the incorporation in recent versions of the public employment chapter of G@G indicators on gender distribution in public employment and the share of women parliamentarians and ministers. These indicators constitute a first attempt to disentangle the characteristics of public employment. The indicators of the shares of women ministers and parliamentarians, which are extracted from the International Parliamentary Union's (IPU) PARLINE database, are considered proxies for the share of women in leadership positions and are important in painting a full picture of women's role in public life.

Levels of remuneration are a key determinant of the capacity of governments to engage talent and thus signal the attractiveness of governments as employers. Remuneration data provide valuable insights into some characteristics of the public sector as evidenced by the great interest in these indicators by the media. However, the absence of information on the different determinants of remuneration (seniority, experience, level of education, gender, etc.) challenges the use of data as key evidence to shape public employment reforms. In addition, a lack of equivalent measures for the private sector makes it difficult to reach any conclusion in terms of the relative situation of public employees when compared to employees in other sectors of the economy.

General assessment: the input indicators in G@G have a **high degree of action worthiness**.

Actionability

The extent to which the input indicators included in G@G can be transformed into concrete actions varies much. On the one hand, the indicators in the public finance and economics chapter are often used as evidence to justify needs for reform. Governments are judged on their fiscal performance and are expected to take action accordingly. The indicators included in G@G are not exclusive, as similar information is generated and displayed by alternative sources both within the OECD and by other international organisations. However, their added value consists in providing a comprehensive picture of public finances rather than scattered indicators.

On the other hand, the indicators in the public employment and remuneration chapter are less actionable. Across the different editions of G@G, the evolution of public employment indicators has shown that public employment tends to remain stable on average across OECD countries over time. Furthermore, even in the context of the global financial and economic crisis, the downsizing measures announced by many governments did not affect the indicators dramatically. A special feature included in the 2015 edition of G@G analyses employment and remuneration reforms that have taken place in OECD countries since 2008. The results of the survey show that 23 of the 29 countries surveyed have implemented remuneration reforms in their central government as a consequence of the financial and economic crisis. However, employment reforms have proven harder to be implemented and downsizing the public sector remains highly contested. Therefore, countries have focused more on less drastic actions such as recruitment freezes, outsourcing, attrition, and voluntary termination.

G@G has incorporated information on the existence of different types of quotas for women parliamentarians. Adding this information endows the indicators with an actionable dimension as the adoption of quotas can be a concrete action to address gender discrimination.

General assessment: the input indicators in G@G are **highly actionable**.

Behavioural

Both in the cases of public finance and public employment, the indicators describe the de facto situation. In practical terms, they are constructed either

as ratios from a head-count exercise (public employment) or by consolidating financial information (public finance), leaving little room for interpretation. The existence of guidelines and manuals, such as the System of National Accounts (SNA) framework, facilitates guaranteeing the relevance of the information. Similarly, the remuneration indicators are mainly quantitative and on many occasions remuneration levels are set by primary or secondary law or collective agreements. In this particular area, there is not much room for practice to differ from the situation stipulated by law.

General assessment: the input indicators in G@G are **very highly behavioural**.

Validity

The input indicators in G@G usually behave in the way theory suggests. For example, all the indicators in the public finance chapter reflected the effects of the global economic crisis, resulting in a deteriorating fiscal stance. Moreover, these indicators are usually sensitive to specific country situations and capture developments in the macroeconomic situation as predicted by macroeconomic theory, hence signalling a significant level of construct validity.

The definitions used for the indicators on public employment are deeply rooted in labour economics theory. Employment and participation rates are widely accepted, used, and understood indicators among policy communities. As previously explained, government employment tends to be relatively stable on average in OECD countries. Therefore, a stable value of the indicator reflects the complexity of putting into practice any downsizing of public administration rather than a failure of the indicator to capture the evolution of the phenomenon under study.

With regard to the remuneration indicator, the underlying remuneration concept is commonly understood. However, the perspective adopted is broader as the indicator endorses the perspective of the cost to government of its employees. This special focus may affect the general understanding of the indicator but it is certainly more effective for governments as it provides a comprehensive picture of the cost of employment and a breakdown on the different components. The lack of variables driving remuneration impedes judgement of the evolution of the indicator according to the assumptions signalled by theory.

General assessment: the input indicators in G@G have a **high degree of validity**.

Reliability

Public finance indicators are calculated on a regular basis (twice a year) following a similar methodology. The existence of the SNA framework guarantees that changes in the value of the indicators mainly reflect the changes in the situation under study and to a lesser extent changes that can be attributed to measurement. Currently, most OECD countries are finalising a migration from the 1993 to the 2008 SNA framework. The change of framework has implied revisiting some of the definitions; in turn this has affected the values of some indicators. However, older data have also been updated according to the new framework in order to maintain a consistent methodology over time.

A similar case can be made for the public employment indicators. While these indicators are collected less frequently than the public finance ones, over time the same methodological approach is followed based on the ILO standards and definitions. Recently, the ILO restricted the scope of its collection efforts, which affected the amount of available information. Given the relevance of this data, collective inter-institutional efforts are being made to replicate the comprehensive data collection exercise. In addition, the employment data by sector (e.g. public or private) produced by the ILO is based on individuals' own assessments of the sector in which they are employed using household surveys. This approach may not be as reliable as collecting data directly from public and private institutions since individuals might not always have a clear understanding of the sector in which they work (for instance, in the case of private but publicly subsidised schools or hospitals).

Remuneration data has been included in G@G on two occasions. However, the two sets of indicators are not comparable as the methodology was changed. The main reason for modifying the methodology was to simplify the questionnaire, which proved extremely complex during the first wave of the survey. As a result, the number of ministries included in the survey was reduced, the categories included within the definitions of remuneration were condensed, and the number of occupational categories analysed was changed.

The main challenge for these indicators is usually definition of government occupations. The OECD follows the International Standard Classification of Occupations (ISCO), from which it has extracted four typical occupational groups in central/federal government. These occupations are considered relatively representative and comparable across countries. Given that not all countries follow the ISCO classification, harmonisation between the classifications is required. On some occasions,

the comparability of the data has been questioned as a result of the harmonisation process and the greater margin granted to some countries to accommodate their employment categories. For future measurement exercises, the objective is to keep the same improved questionnaire in order to be able to compare results over time.

General assessment: the input indicators in G@G have a **high degree of reliability**.

Government Processes

The process indicators in G@G have historically played a dual role. On the one hand, they have been used as evidence to identify the state of the art on a given topic. This evidence has nurtured dialogue between the OECD Secretariat and the policy communities to identify and define best practices. The identification of best practices has led to the definition of policy-oriented tools in the form of recommendations.

On the other hand, the process indicators in G@G have also been used to measure progress in the implementation of the recommendations and principles adopted. In this way, principles often correspond to the theoretical framework for the questionnaire design. By allowing countries to identify their gaps vis-à-vis the best practices, the evidence becomes a key input to identifying policy actions to be implemented in the future.

Since 2010, the OECD council has adopted a series of recommendations in the area of good governance which have guided work on the indicators, including a:

(1) Recommendation on Principles for Transparency and Integrity in Lobbying (2010);
(2) Recommendation on Principles for Public Governance of Public-Private Partnerships (2012);
(3) Recommendation on Fighting Bid Rigging in Public Procurement (2012);
(4) Recommendation on Regulatory Policy and Governance (2012);
(5) Recommendation on Effective Public Investment across Levels of Government (2014);
(6) Recommendation on the Governance of Critical Risk (2014);
(7) Recommendation on Digital Government Strategies (2014);
(8) Recommendation on Budgetary Governance (2015);
(9) Recommendation on Gender Equality in Public Life (2015);
(10) International Open Data Charter (2015).[3]

Action worthiness

In many cases, data collection efforts on public management and governance processes concern a technical domain mainly understandable to a policy community with specific knowledge. In the case of the processes measured in G@G, these policy communities—such as budget officials, HRM experts, and regulatory experts—are directly involved in these measurement efforts.

The public is not necessarily interested in the details of government processes. Often processes are technical and may require an understanding of the legal architecture of the public sector and the specific characteristics of a given sector. Citizens are primarily interested in access to and the quality of the goods and services provided by governments. However, they also care that the services that are provided are those that are needed and that they are delivered efficiently—so that they receive value for their money and the processes used by civil servants are impartial. The action worthiness of government process indicators is greater for policy-makers than for the broader society. General assessment: the process indicators in G@G have **medium/varying degrees of action worthiness**.

Actionability

Process indicators are defined and developed within policy communities and therefore target issues identified as relevant by policy-makers. Originally, and still today, one of the main purposes that triggered the development of G@G was to provide high-quality evidence for policy-making. This evidence would allow reforms to be implemented with the overall objective of improving the quality of governance.

Additionally, a consistent methodology and a permanent discussion forum offer the possibility of benchmarking country performances and learning about how peers are approaching similar challenges. Finally, the pool of available information allows best practices to be identified and recommendations to be issued. The various sets of recommendations provide countries with valuable tools to implement actions according to their specific needs. As discussed earlier, the evidence contained in G@G has been instrumental both as input for the definition of recommendations and as tools to measure progress in their implementation.

General assessment: the process indicators in G@G are **highly actionable**.

Behavioural

One of the main challenges of process data is differentiating between de jure legal mandates and actual de facto practices. While in most cases government officers replying to surveys might have a tendency to respond based on the de jure situation, de facto information provides further insight and more relevant messages for policy communities. Even across OECD member countries, there is a gap between the formal features set out in law and the actual practices of governments. Most of the surveys on government processes focus on legal foundation but they are also increasingly concerned with disentangling the actual practices (de facto).

In order to guarantee that the indicators collected are reflecting observable practices, several steps are taken during the data-cleaning process that takes place once countries have replied to the questionnaires. These entail going back to respondents requiring them to validate their answers. Additionally, in the questionnaires countries are asked to provide evidence in the form of additional documents (e.g. laws, analytical papers, etc.). More recently this procedure has evolved to request countries to provide specific examples of how they are actually implementing what is specified in their answers. The data on government processes collected are gathered every 4–5 years, depending on the specific topic. Consistency over time requires contrasting the answers with those from previous years. While data are not expected to remain constant, changes over time should be explained and documented.

However, despite growing efforts to include behavioural information, the majority of the governance surveys used in G@G represent an institutional mapping of public management practices in terms of the laws, strategies, and directives in place. Recent examples of indicator development in the areas of open data (OURdata Index) and regulatory management (iREG indicators) are shifting away from this focus on purely de jure information.

General assessment: the process indicators in G@G are **behavioural to a medium/varying degree**.

Validity

As mentioned in the introductory section of this chapter, G@G does not produce a unique measure of good governance captured in a super-composite indicator. However, composite indicators are produced for narrowly defined policy areas and in agreement with policy communities. Government processes have proven the most fertile ground for the development of composite

indicators within G@G. In consequence, G@G presents composite indicators on open data, aspects of the budgetary process, disclosure of information by government officers, aspects of human resource management and aspects of regulatory governance.

Figure 9.2 presents an example of a narrowly defined composite indicator in the area of regulatory governance based on the 2012 OECD Recommendation on Regulatory Policy and Governance. This composite indicator measures stakeholders' engagement in developing primary laws in most OECD countries and the European Commission. Stakeholder engagement is defined as the process of communication, consultation, and participation by stakeholders in different phases of the regulatory governance cycle. It measures four main areas: *methodology, systematic adoption, transparency, and oversight and quality control.*[4]

The construction of composite indicators follows a rigorous process, as specified in the OECD guidelines for constructing composite indicators. One of the main challenges in constructing composite indicators lies in guaranteeing that there is agreement on whether or not a given feature is conceptually positive for all the countries to be considered in the index. For example, the OECD has tried to replicate the construction of a composite indicator on budgetary flexibility displayed in the first edition of G@G. However, no

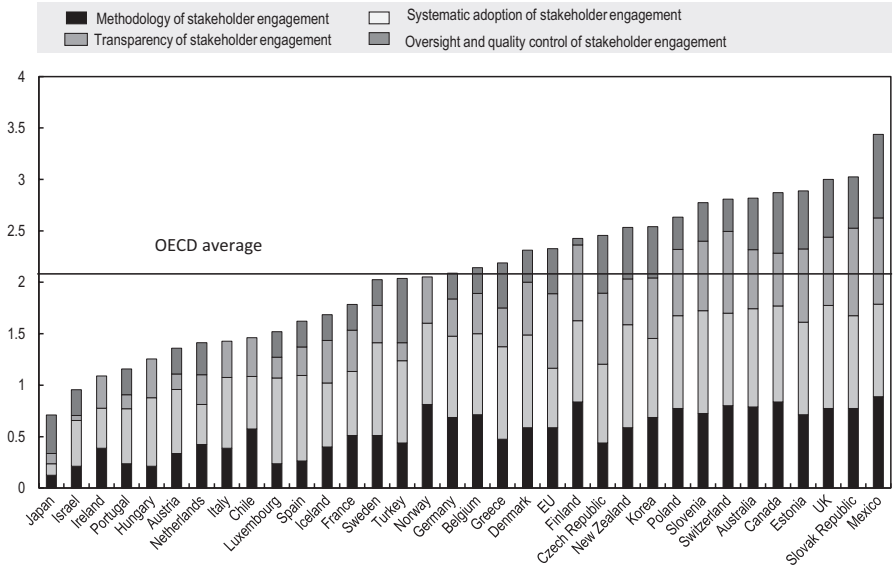

Fig. 9.2 Stakeholder engagement in the development of primary laws, 2015, *scores range from 0 (lowest) to 4 (highest)*. (Source: Data presented in Government at a Glance 2015. Detailed figure notes are accessible in OECD 2015b Regulatory Policy Outlook)

consensus between the member countries has been reached on whether or not more budgetary flexibility is a desirable feature. In consequence, assigning a value based on a composite indicator to whether or not countries have a more or less flexible budgetary process could produce misleading results. Therefore, composite indicators should only be constructed once policy communities have agreed on a direction or practice desirable for all their members.

General assessment: the process indicators in G@G have a **high degree of validity**.

Reliability

The framework for the development of process indicators comprises several steps aimed at guaranteeing the quality of the indicators. It covers all the stages from the development of the questionnaire according to a sound theoretical framework defined in collaboration with policy communities, through the implementation and data-cleaning process, to the final validation of the data. All these steps are taken with the purpose of limiting any possible bias and hence of guaranteeing the reliability of the data. The more extensive and thorough the data-cleaning process is, the higher the degree of reliability that can be achieved.

Recently, an effort has been made to replicate the indicators in G@G to non-member countries, for example, Latin American and Caribbean countries. Consequently, in 2014 the first regional edition of G@G was launched for Latin America and the Caribbean (LAC). Given the absence of permanent working groups in the LAC region, additional steps were taken to guarantee the reliability of the data.

As most of the composite indicators presented in G@G are relatively new, they have only been calculated once, which does not allow time series to be presented. This is also due to the fact that in certain public governance areas, theory and good practices evolve over time and therefore the underlying methodology used to produce the composite indices changes, which does not allow comparison over time. However, in the next G@G 2017, composite indicators in the area of HRM, including the extent of delegation of HR practices and the extent of the use of performance-related pay, will be presented for the first time in a time series, allowing evaluation of the evolution of HRM practices in governments between 2010 and 2016. The objective for future years is to stabilise the composite indicator methodology to the greatest extent possible to allow for more comparison over time.

General assessment: the process indicators in G@G are **highly reliable**.

Government Outputs and Outcomes

While government inputs and internal processes matter in good policy-making, they are mainly assessed in terms of the final results and the impact of policies on the broader society. In G@G, government output and outcome measures are provided in two separate chapters:

(1) **Core government results** (in previous editions: "strategic governance"), focusing on overall government results such as citizen confidence in public institutions, the implementation of rule-of-law principles, fiscal sustainability, fairness in terms of income redistribution, and the government's ability to carry out its tasks in an efficient and cost-effective way.
(2) **Serving citizens** (sectoral outputs and outcomes), assessing the levels of access, responsiveness, and quality of public service delivery to citizens in areas such as healthcare, education, justice, and tax administration (Table 9.3).

For these two chapters, the OECD works closely with many other international institutions and, internally, has developed various forms of inter-directorate cooperation.[5] The great majority of the data provided in this section are individual indicators using a mix of subjective and administrative data.

The main challenge related to these output and outcome measures lies in isolating the effect of government action on observable results from the effect of other external factors (attribution problem). In addition, the currently available data for many countries do not always clearly separate out the performance of public and private institutions (for instance, hospitals and schools), which limits the possibility for comparative analyses between the public and private sector. This represents an important data and research agenda that will need to be pursued in the coming years.

Table 9.3 OECD framework for measuring services to citizens, 2015

Access	Responsiveness	Reliability/quality
Affordability	Citizen-centred approach (courtesy, treatment, and integrated services)	Effective delivery of services and outcomes
Geographic proximity	Match of services to special needs	Consistency in service delivery and outcomes
Accessibility of information	Timeliness	Security (safety)

Source: OECD Government at a Glance (2015a)

Action worthiness

There is wide agreement in the literature that output and outcome indicators are fundamental components of measures of government performance (Rotberg 2014; Boardman 2014; Holt and Manning 2014; Fukuyama 2013).

Core Government Results These indicators are very important since they represent key strategic outcomes of government activities. They are strategic in the sense that they are crucial to the well-being of societies and economies. Many of them reflect the performance of government functions that cannot be carried out by any other actor, for example, income redistribution, rule of law, and risk management. Increasing strategic foresight, fairness and equity, efficiency, cost-effectiveness, and levels of confidence in public institutions are generally key objectives of governments and are reflected in the national public sector reform documents of many OECD countries.

Serving Citizens Improving services to citizens is also a key objective of public sector reforms in OECD countries. Governments are responsible for providing a wide range of public services that should meet the expectations of their citizens in terms of access, responsiveness, and quality. The direct experience of citizens of front-line public services in healthcare, education, transport, justice, employment, tax administration, and other areas greatly affects their satisfaction with these services and, more broadly, their trust in public institutions.

General assessment: the output and outcome indicators in G@G have a **very high degree of action worthiness**.

Actionability

Core Government Results The degree to which the indicators provided in the strategic outcome chapter of G@G are actionable for governments varies greatly. Most of the levels of the indicators provided in this section and the changes in them do not depend solely on government actions and may be hard to take action on. For instance, the level of a population's confidence in the national government collected by the Gallup World Poll depends on a wide range of factors on which governments may have little control (e.g. media campaigns). In addition, the cost-of-collection ratio measuring the efficiency of tax administration (tax administration costs/total tax revenues) is strongly dependent, on the

denominator side, on the macroeconomic conditions which affect tax receipts and on the numerator side on the type of fiscal system in place.

Serving Citizens By comparison, the indicators provided on the levels of access, responsiveness, and quality of services are more actionable for governments. For instance, to address geographical access to healthcare, different types of financial incentives are offered to doctors to attract them to underserved areas and retain them there, including one-off subsidies to help them set up their practices and recurrent payments such as income guarantees and bonus payments (OECD 2013b). Similarly, to address long waiting times for appointments with specialists or doctors, waiting time guarantees have become the most common policy tool in several OECD countries over the last decade (Siciliani et al. 2013; OECD 2013c). In the area of justice too, many governments have equipped judicial courts with electronic systems and now allow for the submission and follow-up of certain claims online, which can have a direct impact on the length and timeliness of judicial procedures (CEPEJ 2014).

General assessment: the output/outcome indicators in G@G are **actionable to a medium/varying degree**.

Behavioural

The majority of the indicators provided in the G@G core government results chapter and the serving citizens chapter reflect observable facts and implementation (rather than rules, directives, and institutional bodies/settings) and as such are generally highly behavioural. For instance, the expert and household survey questions used in the World Justice Project's Rule of Law (RoL) Index are very highly behavioural. They ask questions based on hypothetical scenarios to assess the experience of citizens of the rule of law and evaluate expert judgements (World Justice Project 2014).[6]

Similarly, in the serving citizens chapter, both the subjective data and the administrative data generally focus on observable behaviour. For instance, the indicators used to measure citizen access to legal information are based on the European Commission's Special Eurobarometers, where respondents are asked to assess how informed they feel about legal procedures in case they have to go to court (finding a lawyer, cost of procedures, etc.). The quality of education and equity in learning outcomes are measured using the scores obtained by students in the 2012 Programme for International Student Assessment (PISA), which is an exercise-based assessment of the competencies of 15-year-olds in reading, mathematics, and science in 65 countries (OECD 2013d).

General assessment: the output/outcome indicators in G@G are **very highly behavioural**.

Validity

Core Government Results The degree of validity of the indicators provided in this chapter varies greatly. On the one hand, the RoL Index was audited for the fourth time in 2014 by the Econometrics and Applied Statistics Unit at the European Commission Joint Research Centre (JRC), which again confirmed that the index is overall statistically sound in terms of coherence and balance. The assessments of the conceptual and statistical coherence of the index and that of the impact of modelling choices on country performances contribute to building stronger "confidence in the narrative supported by the measure" (Saisana and Saltelli 2014). The assessment of the robustness of the RoL Index was based on a combination of a Monte Carlo experiment and a multi-modelling approach. The auditing exercise concluded that the RoL Index is statistically "robust" without being redundant and this despite very high correlations (>95) between some of the factors.

On the other hand, some other output and outcome measures used in the chapter might face some validity issues. For instance, preliminary evidence suggests that the measures of confidence in the national government from various large-scale household surveys face a significant "convergence" validity issue since their correlation with other existing measures of the same underlying concept is rather low (despite similar representative sample sizes). The OECD has developed *Trust Lab* experiments to improve interpersonal and institutional trust measures, and they are currently being tested in France and Korea.

Serving Citizens Most of the indicators provided in the serving citizens chapter represent valid measures of the underlying concept. In most cases, they are taken from the administrative records of hospitals, schools, and courts. For the OECD indicators (which represent the majority of the indicators in this section), despite some persisting definition and conceptual issues, generally for specific countries, these indicators are considered to be among the most valid existing measures at the international level. They have also benefited from extensive work in committees to harmonise definitions and methodologies, for instance, in the case of PISA or as part of the Health Care Quality Indicators Project (see for example Kelley and Hurst 2006).

General assessment: the output/outcome indicators in G@G have a **medium/varying degree of validity**.

Reliability

Core Government Results The degree of reliability of the indicators provided in this chapter varies greatly. Most of the indicators from administrative sources (e.g. life expectancy, cost of tax collection, expenditure on healthcare, expenditure on education, and average length of stay in hospitals) are generally reliable measures based on common/harmonised methodologies.

However, the data from household surveys and expert surveys used in G@G are often not as reliable across populations and over time, due mainly to poorly defined concepts and questions and limited sample sizes in most cases. Cultural factors may influence the responses to subjective survey questions due to differences in expectations and also different interpretations of semantics. Moreover, contextual elements and media campaigns can often affect scores despite the underlying concept/situation being measured as not having evolved in reality.

Serving Citizens In general, the indicators provided in the serving citizens chapter, which mainly come from harmonised tests (PISA) and administrative data collection, are generally more reliable across populations and over time. This is due to significant efforts to harmonise concepts and definitions across countries by OECD working parties and expert groups to ensure reliability/comparability of the information provided.

General assessment: the output and outcome indicators in G@G are **highly reliable**.

Conclusion

Overall, informing public sector reforms in OECD countries by means of robust and actionable evidence requires a strong focus on conceptualisation and a narrowly defined approach to measuring government activities. This is what G@G has been doing through close collaboration with member countries over the past 7 years using data collected from government officials, National Statistical Offices, and other organisations, breaking down the concept of "good governance" into more measurable elements. Even though not all the indicators provided in the publication fully meet the assessment criteria in the framework provided in this chapter, the OECD is working in collaboration with its member countries to improve the level of relevance and robustness of the indicator set in every edition of G@G.

To effectively inform public decision-making, a dashboard approach to indicator building has been followed. This is reflected in the construction of a set of narrowly defined composite indices on public management practices (e.g. performance budgeting, asset disclosure, regulatory management, human resource management) where quantitative data is missing or they do not constitute appropriate measures based on OECD instruments setting the principles and recommendations on commonly agreed good practices. A mix between household opinion survey data, expert survey data, and administrative data is also provided in the publication for triangulation purposes, taking into account the difficulty in accurately interpreting data from different sources and considering their different strengths and weaknesses.

The G@G dashboard approach allows discussion to move away from aggregation methods and arbitrary weighting decisions to focus instead on conceptualisation, definitions, and substantive elements of indicator building in close collaboration with government officials. As such, some of the indicators provided in the publication might not be as "media" friendly as the super-composites developed by other organisations. This could explain, at least partly, the relative lack of knowledge of the publication among the wider population and the academic world. In recent years, stronger efforts have been made to advertise the G@G indicators and to explain to a wider audience how good-quality indicators can inform government reforms and improve their effectiveness.

While the G@G publication provides evidence, it does not measure trade-offs between different policy options. This is deliberately left to policy-makers and other OECD publications which analyse the timing of reforms in national jurisdictions and reform scenarios. It is also left to the broader community of researchers who, starting from 2015, can access all the data online at no cost. The strong emphasis on transparency allows for more in-depth analyses by external stakeholders, greater re-use of data, and an increase in the accuracy of the data provided in the publication.

Notes

1. Administrative data are data collected from the records of administrative units (e.g. hospital records, school records, court records, etc.). Administrative data are often considered to be of higher quality because they are based on tangible records rather than on perceptions or expert opinions. As such, they represent objective "facts" rather than subjective "interpretations of the facts".
2. Individual (or single) indicators measure precise confined concepts which generally correspond to clearly defined accounting methodologies (such as the System of National Accounts for public finance data) and contain the responses

to specific questions (qualitative or quantitative) in the household and expert surveys.

3. This was not an OECD recommendation, but the OECD Public Governance Directorate contributed to its drafting. The Charter was used in the work on indicators carried out by the OECD digital government unit in collaboration with country delegates, http://www.oecd.org/gov/digital-government/.
4. "Methodology" examines the existence of guidance documents, methods, and tools used for stakeholder engagement, including minimum periods for consultation and the use of interactive websites and social media tools. "Systematic adoption" investigates whether there are formal requirements for stakeholder engagement and the extent to which stakeholders are engaged in practice both in the early and in the later stages of the regulation-making process. "Transparency" looks at the extent to which the processes of stakeholder engagement are open to the widest spectrum of stakeholders and whether and how stakeholders' views and comments are taken into account. "Oversight and quality control" measures whether there are mechanisms in place to externally control the quality of stakeholder engagement practices (mostly public consultations) and monitor stakeholder engagement and whether evaluations are made available publicly.
5. In the core government results chapter, the indicators come from: the Gallup World Poll for the data on satisfaction and confidence in public services and institutions, the World Justice Project for the data on the rule of law, and from the OECD Health Directorate and Education Directorate for the efficiency and cost-effectiveness indicators focusing on these two sectors. Regarding sectoral outputs and outcomes, the data mainly come from the OECD Health Directorate, the OECD Education Directorate, and the European Commission for the Efficiency of Justice (CEPEJ).
6. For example, the household survey asks respondents to answer the following question: "Please assume that the government decides to build a major public works project in your neighbourhood (such as a railway station or a highway). How likely are people in your neighbourhood/members of your community to be given the opportunity to express their opinions on the project?" The scale is from 1 (very likely) to 4 (very unlikely).

References

Arndt, C. (2008). The Politics of Governance Ratings. *International Public Management Journal, 11*(3), 275–297.

Arndt, C., & Oman, C. (2006). *Use and Abuse of Governance Indicators*. Paris: OECD, Development Centre Studies.

Carmines, E. G., & Zeller, R. A. (1979). *Reliability and Validity Assessment*. Beverly Hills: Sage Publications.

CEPEJ. (2014). *Evaluation of European Judicial Systems* (5th Report). Strasbourg: Council of Europe Publishing.
Fukuyama, F. (2013). *What Is Good Governance* (Governance review. Working Paper 314, January). Washington, DC: Center for Global Development.
Holt, J., & Nick, M. (2014). Fukuyama Is Right About Measuring State Quality: Now What? *Governance Review, 27*(4), 717–728.
Hood, C. (1991). A Public Management for All Seasons? *Public Administration, 69*(1), 3–19.
Kaufmann, D., & Kraay, A. (2007). *On Measuring Governance: Framing Issues for Debate* (Issues Paper for January 11th, 2007 Roundtable on Measuring Governance). Washington, DC: World Bank Institute.
Kelley, E., & Hurst, J. (2006). *Health Care Quality Indicators Project: Conceptual Framework Paper* (OECD Health Working Papers, No. 23). Paris: OECD Publishing. Retrieved from http://dx.doi.org/10.1787/440134737301
Nardo, M., et al. (2005). *Handbook on Constructing Composite Indicators: Methodology and User Guide* (OECD Statistics Working Papers, No. 2005/03). Paris: OECD Publishing.
OECD. (2009). *Government at a Glance 2009*. Paris: OECD Publishing.
OECD. (2011). *Government at a Glance 2011*. Paris: OECD Publishing.
OECD. (2012a). *The OECD Quality Framework for Statistical Activities*. Paris: OECD Statistics Directorate.
OECD. (2012b). *Measuring Regulatory Performance: A Practitioner's Guide to Perception Surveys*. Paris: OECD Publishing.
OECD. (2013a). *Government at a Glance 2013*. Paris: OECD Publishing.
OECD. (2013b). *Regions at a Glance 2013*. Paris: OECD Publishing.
OECD. (2013c). *Health at a Glance 2013: OECD Indicators*. Paris: OECD Publishing.
OECD. (2013d). *PISA 2012 Results: Excellence Through Equity: Giving Every Student the Chance to Succeed* (Vol. II). Paris: OECD Publishing.
OECD. (2014). *OECD Framework for Regulatory Policy Evaluation*. Paris: OECD Publishing.
OECD. (2015a). *Government at a Glance 2015*. Paris: OECD Publishing.
OECD. (2015b). *OECD Regulatory Policy Outlook 2015*. Paris: OECD Publishing.
Osborne, D., & Gaebler, T. (1992). *Reinventing Government: How the Entrepreneurial Spirit is Transforming the Public Sector* (427pp). Reading: Addison-Wesley Publishing.
Rotberg, R. I. (2014). Good Governance Means Performance and Results. *Governance: An International Journal of Policy, Administration, and Institutions, 27*(3), 511–518.
Saisana, M., & Saltelli, A. (2014*). JRC Statistical Audit of the WJP Rule of Law Index 2014* (EC Joint Research Centre, Working Paper for the 2014 Rule of Law Index).

Saisana, M., & Tarantola, S. (2002). *State-of-the-Art Report on Current Methodologies and Practices for Composite Indicator Development, EUR 20408*. Ispra: European Commission Joint Research Center, Institute for the Protection and the Security of the Citizen. Technological and Economic Risk Management Unit.

Siciliani, L., Borowitz, M., & Moran, V. (2013). *Waiting Time Policies in the Health Sector: What Works? OECD Health Policy Studies*. Paris: OECD Publishing.

World Justice Project. (2014). *The Rule of Law Index 2014 report*. Washington, DC: World Justice Project.

10

The European Commission's Fiscal Sustainability Indicators and Their Use in the EU's Integrated Cycle of Economic Policy Coordination

Katia Berti

Introduction

Fiscal sustainability is one of the core areas of surveillance for the European Commission's Directorate General for Economic and Financial Affairs (DG ECFIN). At the centre of its analysis is the solvency of the public sector, defined as its ability to honour current and future financial obligations on time, including the implicit commitment to continue providing certain public goods, services and transfers.

The analytical framework used to support the European Commission's fiscal sustainability assessments relies on three indicators (S0, S1 and S2) which are designed to capture fiscal sustainability challenges over the short, medium and long term. The role of the indicators is that of a toolkit to help the reading and understanding of the budgetary situation and the potential fiscal risks for the country under examination, additionally taking into account projected implicit liabilities for the government related to population ageing. In this spirit, categorisation as low, medium or high risk in the short, medium and long term according to the three sustainability indicators serves the purpose of providing a consistent horizontal analysis across countries of the specific dimension of fiscal sustainability. The framework is meant to identify risks to

Figures in this chapter have been originally published in the European Commission, 2015, Fiscal Sustainability Report 2015, http://dx.doi.org/10.2765/412671.

K. Berti (✉)
European Commission, Brussel, Belgium

D.V. Malito et al. (eds.), *The Palgrave Handbook of Indicators in Global Governance*,
https://doi.org/10.1007/978-3-319-62707-6_10

fiscal sustainability early on and also to qualify risks by explicitly considering the time dimension over which they could materialise. Both features are key to supporting timely and appropriate policy responses.

After an analysis of the genesis of the fiscal sustainability indicators used by the Commission services and their definitions (as from the latest European Commission's Fiscal Sustainability Report 2015a), this chapter provides a concrete example of the results published in the context of the European Semester[1] 2014 and explains in detail how the indicators are used in the policy process.

Genesis and Current Design of the European Commission's Fiscal Sustainability Indicators

The European Commission's fiscal sustainability indicators have been developed over time with successive changes and integrations. These were results of efforts aimed at making the assessment framework as comprehensive as possible and better able to capture additional challenges to fiscal sustainability identified along the way.

Originally, the framework only relied on so-called sustainability (or tax) gap indicators (S1 and S2) and focussed exclusively on the long term (European Commission 2006). This is because the framework was conceived to capture the budgetary effects of implicit government liabilities related to population ageing, which are meaningfully covered with a long time horizon. Sustainability gap indicators like S1 and S2 provide measures of the additional fiscal adjustment effort (in structural terms) needed to ensure fiscal sustainability (in the sense of solvency of the public sector, as explained above). More precisely, for the S2 indicator, fiscal sustainability translates into the requirement that public debt as a percentage of GDP does not follow an explosive path over the infinite time horizon when also taking projected public spending related to population ageing (pensions, healthcare and long-term care) into account.[2] This definition has remained unchanged since the European Commission's original Fiscal Sustainability Report 2006. The S2 indicator, as just described, is indeed still the tool currently in use to detect long-term fiscal sustainability challenges posed by the costs of ageing (European Commission 2015a).

The definition of the S1 indicator has, on the contrary, changed over time. The main objective pursued by the European Commission with the introduction of the S1 indicator was to have a measure of fiscal sustainability challenges based on a concept of fiscal sustainability that would better reflect EU fiscal rules by incorporating the 60% Stability and Growth Pact reference

value for public debt over GDP. In the 2006 Fiscal Sustainability Report, the S1 indicator was indeed defined as a measure of the fiscal adjustment effort needed to ensure the public debt target of 60% of GDP would be reached in 2050 taking account of the projected increase in public spending due to population ageing up to that year. The difference in this definition relative to the S2 indicator lies in the reference to a finite (rather than infinite) time horizon (thought of as more relevant from a policy perspective), and in the additional requirement of a specific level for the debt-to-GDP ratio (in line with EU fiscal rules) to be reached within the finite horizon. In the subsequent Fiscal Sustainability Report 2009, the definition of the S1 indicator was kept unchanged with the exception that the year for reaching the debt ratio target was moved back to 2060 (European Commission 2009). However, a more structural change has recently been introduced in the definition of the indicator. This is because the need was felt to assign a more distinctive identity to the S1 indicator and translate it into an indicator of medium-term sustainability challenges, thus further differentiating its time dimension from the long-term perspective adopted with the S2 indicator. In the 2012 Fiscal Sustainability Report, the S1 indicator was therefore re-defined as a measure of the cumulative fiscal adjustment effort (in structural terms) to be made by 2020 to reach the public debt target of 60% of GDP by 2030. The definition of the indicator was then further refined in the last release of the Fiscal Sustainability Report (2015a) to describe the cumulative structural adjustment to be done over 5 years to reach the 60% debt target by 2030 (European Commission 2015a).

A very appealing feature of both the S1 and S2 indicators, from the point of view of the analytical support provided to policy advice, consists in the fact that the two can be expressed as the sum of different components respectively referring to the country's fiscal stance and to the projected budgetary effects of population ageing (see Figure 10.1). In S2, the component related to the fiscal stance consists exclusively of the so-called initial budgetary position (determined by the government's primary balance in structural terms and the public debt over GDP in the last forecast year preceding the start of the projections). This component importantly provides a measure of the existing gap relative to the debt-stabilising primary balance (the structural primary balance that would stabilise the public debt-to-GDP ratio given the initial debt ratio, the interest rate and the growth rate). In the S1 indicator, the component linked to the fiscal stance is more complex in that it incorporates two sub-components: the initial budgetary position (analogous to that in the S2 indicator)[3] and the required additional fiscal adjustment needed to reach the 60% debt-ratio target (the latter is determined by the distance of the debt-to-GDP ratio from

	Required adjustment given initial budgetary position		Required adjustment to reach debt ratio target of 60% in 2030		Required adjustment due to cost of ageing
S1 =	Gap to debt-stabilising primary balance	+	Additional adjustment required to reach 60% debt ratio in 2030	+	Additional adjustment required to finance the increase in public spending due to ageing up to 2030
S2 =	Gap to debt-stabilising primary balance	+	0	+	Additional adjustment required to finance the increase in public spending due to ageing over infinite horizon

Fig. 10.1 The S1 and S2 indicators and their components. Source: European Commission (2015a)

the 60% target at the start of the projections). Finally, both the S1 and S2 indicators include a component that relates to the additional fiscal adjustment required by the projected public spending due to ageing. The latter can be usefully further disaggregated into the individual components that make up ageing costs, distinguishing between public spending on pensions, healthcare and long-term care. This allows examination of the individual determinants that are behind the projected budgetary impact of ageing in a way that is most useful to provide policy advice and formulate policy recommendations to Member States, as better explained in section "The Use of the European Commission's Fiscal Sustainability Indicators in the Policy Process".

While the (modified, as explained above) S1 indicator and the S2 indicator already allowed challenges related to countries' budgetary positions and projected ageing costs over the medium to long term to be captured, the economic and financial crisis, which later turned into a sovereign debt crisis, highlighted the need to complement the Commission's fiscal sustainability analysis framework in two respects: (a) coverage of short-term challenges (by definition left out in the traditional S1 and S2 indicators) and (b) coverage of possible fiscal risks stemming from the macro-financial side of the economy beyond those stemming directly from the fiscal side. Both points were tackled with the development of a new indicator (S0), which, together with the S1 and S2 indicators, is now an integral part of the Commission's fiscal sustainability analysis framework (see European Commission 2015a, 2012).

The S0 indicator is based on a different methodological approach relative to the other two indicators. S0 does not belong to the category of sustainability gap indicators but is a composite indicator constructed using 28 variables (14 fiscal variables and 14 variables related to macro-financial and competitiveness factors—see Table 10.1) and is based on the so-called signals

Table 10.1 The S0 indicator, its fiscal and financial competitiveness sub-indexes and the individual variables included thresholds and signalling power

Variables	Safety	Threshold	Signalling power	Type I error	Type II error
Balance, % GDP	>	−10.17	0.07	0.04	0.89
Primary balance, % GDP	>	0.00	0.17	0.40	0.43
Cyclically adjusted balance, % GDP	>	−3.12	0.25	0.45	0.30
Stabilising primary balance, % GDP	<	2.55	0.02	0.12	0.86
Gross debt, % GDP	<	103.28	0.03	0.06	0.91
Change in gross debt, % GDP	<	6.50	0.11	0.08	0.81
Short-term debt, government, % GDP	<	16.00	0.10	0.11	0.79
Net debt, % GDP	<	58.11	0.13	0.19	0.68
Gross financing needs, % GDP	<	16.83	0.16	0.21	0.63
Interest rate-growth rate differential	<	5.92	0.08	0.07	0.85
Change in expenditure of gen. government, % GDP	<	2.25	0.14	0.13	0.74
Change in final consumption expend. of gen. government, % GDP	<	0.64	0.17	0.19	0.64
Old-age dependency ratio 20 years ahead	<	33.93	0.10	0.11	0.79
Avg. yearly change in projected age-related public expend. as % of GDP over next 5 years	<	0.26	0.09	0.14	0.77
Fiscal index	<	0.35	0.23	0.21	0.56
L1.net international investment position, % GDP	>	−50.10	0.31	0.13	0.56
L1.net savings of households, % GDP	>	0.96	0.34	0.26	0.40
L1.private sector debt, % GDP	<	209.20	0.25	0.04	0.71
L1.private sector credit flow, % GDP	<	10.90	0.44	0.42	0.14
L1.leverage, financial corporations	<	2.22	0.03	0.97	0.00
L1.short-term debt, non-financial corporations, % GDP	<	27.40	0.25	0.21	0.54
L1.short-term debt, households, % GDP	<	3.50	0.27	0.34	0.38
L1.construction, % value added	<	7.25	0.27	0.36	0.38
L1.current account, 3-year backward MA, % GDP	>	−2.45	0.38	0.37	0.25
L1.change (3 years) of real eff. exchange rate, based on exports deflator	<	9.76	0.23	0.19	0.59
L1.change (3 years) in nominal unit labour costs	<	12.70	0.27	0.48	0.25
Yield curve	>	0.59	0.48	0.39	0.14
Real GDP growth	>	−0.89	0.10	0.07	0.83
GDP per capita in PPP, % of US level	>	73.32	0.28	0.44	0.27
Financial competitiveness index	<	0.45	0.48	0.34	0.18
Overall index	<	0.43	0.55	0.21	0.25

Source: European Commission (2015a)

approach (see European Commission 2012; Berti et al. 2012; Kaminsky, Lizondo and Reinhart 1998; Kaminsky and Reinhart 1999). The variables included in S0 are chosen based on the economic literature on early warning systems and on analysis of the performance of the variables in detecting past situations of short-term fiscal stress. Most of the variables included in the scoreboard for the surveillance of macroeconomic imbalances (used in the context of the Macroeconomic Imbalances Procedure—see European Commission 2015b) are among the financial competitiveness variables included in S0. Being also based on the most recent experience in the EU, this duly reflects the evidence of the role that financial and competitiveness variables can play in generating potential fiscal risks.

The S0 indicator can be defined as an "early-detection indicator" designed to highlight short-term risks of fiscal stress with a 1-year horizon (Baldacci et al. 2011)[4] stemming from both the fiscal and the macro-financial and competitiveness sides of the economy (European Commission 2015a, 2012; Berti et al. 2012). The signals approach—the methodology used for the S0 indicator—allows an endogenous determination of the thresholds of fiscal risk for the composite indicator itself, for each individual variable included and for the two thematic sub-indexes, respectively, incorporating only fiscal and financial competitiveness variables (for the thresholds, as from the Fiscal Sustainability Report 2015a, see Table 10.1).[5] Values beyond the respective thresholds of the S0 indicator, the individual variables and the two sub-indexes are read as signals of potential upcoming short-term fiscal risks. In particular, for the overall composite indicator S0, a value above the threshold signals potential short-term risks of fiscal stress, as by construction the value of the indicator is greater the higher the number of variables signalling fiscal risks and the better their historical record at highlighting fiscal risks (their "signalling power"—see Table 10.1) (see European Commission 2015a; Berti et al. 2012).

Overall, shorter-term fiscal sustainability challenges can be assessed by focussing on the value of the S0 indicator, while its two thematic sub-indexes further allow risks to be identified emanating from specific areas (fiscal or financial competitiveness) that may or may not translate into fiscal risks signalled by the overall S0 indicator. For countries for which fiscal risks emerge with regard to one of the two sub-groups of variables if the S0 signals no risks, short-term challenges (which arise with regard to either the fiscal or the financial competitiveness side of the economy) are not so acute as to generate risks of fiscal stress at the aggregate level. Consideration of the two sub-indexes is, moreover, relevant to gain insights into the specific area(s) from which risks stem for countries for which overall fiscal sustainability risks are found to be

high according to S0. A more precise identification of the specific sources of the short-term risk of fiscal stress at the country level is supported by analysis of the individual variables and the values they take relative to their own thresholds. Going down to the level of the S0 sub-indexes and the individual variables allows identification of the determinants of possible fiscal risks, which is clearly the first step to support the devising of policy actions.

A comprehensive assessment of fiscal sustainability challenges faced by Member States over the short, medium and long term relies on joint consideration of the three indicators, S0, S1 and S2. Following the presentation of the framework for the first time in the Commission's Fiscal Sustainability Report 2012, the ECOFIN Council welcomed this "multi-dimensional approach" to fiscal sustainability analysis (see Council of the European Union 2013).

The European Commission's Results from Fiscal Sustainability Analysis in the Context of the 2014 European Semester

To make the description of the European Commission's fiscal sustainability assessment framework more concrete, this section briefly presents the results published in the Commission's 2014 Staff Working Documents (an explanation of the institutional processes in which fiscal sustainability indicators are used follows in the next section, see European Commission 2014c). These documents, produced in the context of the Commission's reporting for the European Semester, represented the vehicle through which the values of the fiscal sustainability indicators were presented to the public once a year. More recently, since the European Semester 2015, this role has been taken up by the Commission Assessment of the Stability and Convergence Programmes, published in spring each year, where values of the sustainability indicators are systematically reported. Although reporting is annual, the indicators are nonetheless updated more frequently (three times a year, following the release of the Commission's macroeconomic forecasts) to support regular analysis conducted within the Directorate-General for Economic and Financial Affairs (DG ECFIN) and to keep EU Member State delegates in the Economic and Financial Committee informed about developments concerning fiscal sustainability challenges.

Figures 10.2 and 10.3 provide the measures of the fiscal sustainability challenges faced by Member States over the medium and long terms based on the S1 and S2 indicators published for the European Semester 2014.[6] The charts report both the overall values of the indicators and the components that allow

identification of the specific areas that challenges stem from (the fiscal stance of the country under examination at the start of the projections and the projected dynamics of its age-related public spending on pensions, healthcare and long-term care, as explained in the previous section).

Figure 10.2, for instance, shows that for the EU as a whole, the medium-term fiscal sustainability gap (based on the European Commission's spring 2014 forecasts) was 1.7% of GDP. This means that the required improvement in the primary balance (in structural terms) to achieve a public debt target of 60% of GDP by 2030 for the EU amounted to 1.7% of GDP over the post-forecast period 2016–2020[7] (an average budgetary consolidation effort of approximately 0.3% per year). In other words, the EU's average structural primary balance would have to improve from a projected surplus of 1.1% of GDP in 2015 (based on spring 2014 forecasts) to a surplus of 2.8% in 2020. A medium-term fiscal sustainability challenge of this magnitude is classified as a "medium risk" in the Commission's fiscal sustainability assessment framework.[8]

The public debt target of 60% of GDP by 2030 is the element that contributed the most to the required fiscal consolidation effort highlighted by the S1 indicator for the EU for the 2014 European Semester (see Fig. 10.2). The projected budgetary effects of ageing also contributed to raising the EU's required fiscal adjustment, but to a much smaller extent. On the contrary, starting from a projected structural primary surplus of 1.1% of GDP in the last forecast year, 2015, the initial budgetary position at aggregate EU level contributed to a reduction in the required adjustment. As shown in Fig. 10.2, behind the EU aggregate significant differences were recorded across the

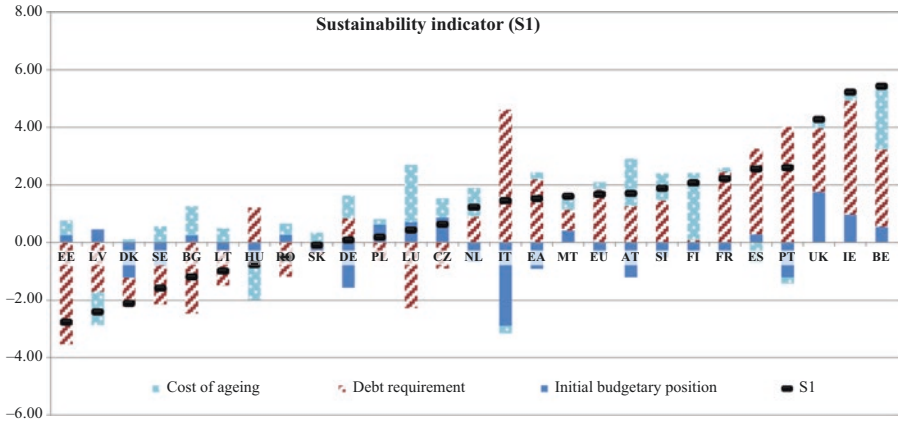

Fig. 10.2 The S1 indicator and its components. Source: Own design based on data from European Commission (Staff Working Documents, June 2014c)

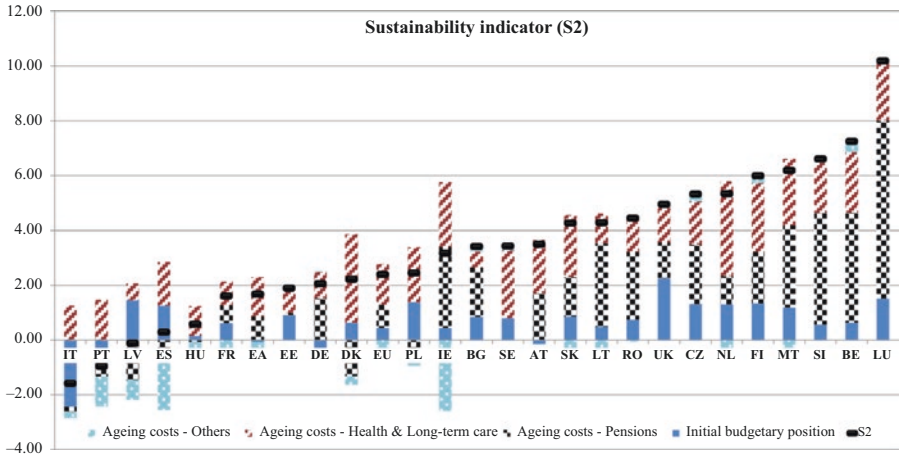

Fig. 10.3 The S2 indicator and its components. Source: Own design based on data from European Commission (Staff Working Documents, June 2014c)

Member States, both in terms of the overall required fiscal adjustment and the underlying determinants of the fiscal sustainability gap.

In terms of long-term fiscal sustainability challenges, an S2 indicator of 2.4% of GDP was reported for the EU as a whole (based on the European Commission's spring 2014 forecasts), also highlighting a "medium risk" in the longer term (see Fig. 10.3).[9] The adjustment required due to the cost of ageing was the main component (2.0% of GDP) behind the EU sustainability gap. Within the cost of ageing, healthcare and long-term care expenditure were the main drivers of the adjustment required (1.5% of GDP), followed by pension expenditure (0.8%). Also in this case, significant cross-country differences in the overall extent of the challenges and in the factors determining them were reported, as is evident from Fig. 10.3.

Finally, shorter-term fiscal sustainability challenges based on the S0 indicator are reported by country in Fig. 10.4. Along this time dimension too, big differences were evident across the countries in spring 2014. At the same time, it also emerges that the shorter-term fiscal stress risks for 2014 (based on 2013 data) had significantly abated for all the countries relative to 2009. Based on the update of the indicator following the Commission's spring 2014 forecasts, only Portugal appeared to be at short-term risk of fiscal stress in 2014 (with an S0 value of 0.53, above the critical threshold of 0.43). This was due to both the S0 fiscal and financial competitiveness sub-indexes being above their thresholds. As already explained in the previous section, analysis of the individual variables included in S0 (whether they take values above or below their

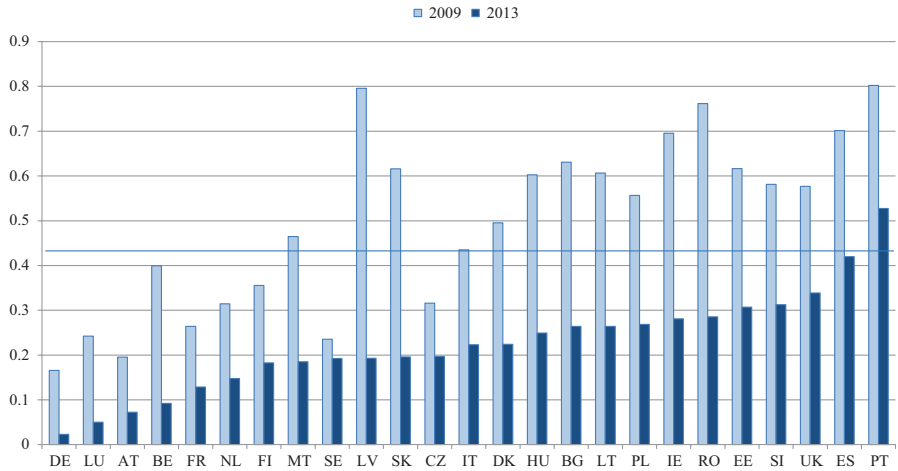

Fig. 10.4 The S0 indicator. Source: Own design based on data from European Commission (Staff Working Documents, June 2014c) and European Commission (2012)

critical risk thresholds) further allows identification of the determinants of the risk signal sent by the S0 indicator, which is clearly particularly important to support the devising of appropriate policy responses.

The Use of the European Commission's Fiscal Sustainability Indicators in the Policy Process

The fiscal sustainability indicators described in this chapter are used to support policy advice and recommendations to Member States in the context of the European Semester, which is the EU's integrated annual cycle of economic and budgetary policy coordination (see European Commission 2014b). In the European Semester, guidance is provided to Member States (before they take policy decisions at the national level) with regard to: (a) the coordination of fiscal policies under the Stability and Growth Pact, (b) macroeconomic policy under surveillance in the context of the Macroeconomic Imbalances Procedure and (c) the implementation of structural policies in line with the Europe 2020 strategy (the EU's strategy for growth and jobs).

The European Semester starts with the Commission publishing the Annual Growth Survey (AGS) in November each year. This is aimed at outlining the broad economic policy priorities for the EU (see European Commission 2015c).[10] National programmes are presented in the spring by

the Member States. Based on assessment of each country's economic situation and the national programme submitted, in May each year the European Commission makes country-specific recommendations (CSRs) to each Member State covering fiscal and macroeconomic policy together with structural reforms. The recommendations proposed by the Commission are then discussed in the Council and endorsed by the European Council, before being finally adopted. The Member States are then expected to reflect the recommendations in their budgetary and policy plans for the subsequent year and to implement them in the following 12 months (European Commission 2014b).

The European Commission's CSRs are accompanied by the Commission's Staff Working Documents (SWDs), one per Member State. The aim of these documents is to provide underpinnings to the CSRs by concisely presenting the analysis that supports the recommendations proposed by the Commission (2014b).[11]

It is in this context that the Commission's fiscal sustainability indicators are used to provide analytical input and support the formulation of the Commission's proposed CSRs. The SWDs are indeed based, among other things, on the identification of short-, medium- and long-term challenges highlighted by the S0, S1 and S2 indicators.[12] The role of the fiscal sustainability indicators is thus that of a toolkit to help the reading and understanding of the budgetary situation and the potential fiscal risks for the country under examination, additionally taking into account projected implicit liabilities related to population ageing. In this spirit, categorisation as low, medium or high risk over the short, medium and long term based on the three fiscal sustainability indicators serves the purpose of providing a consistent horizontal analysis across countries along the specific dimension of fiscal sustainability. This is then further qualified considering country specificities and integrated into the wider analysis of economic and social challenges to the countries conducted by the Commission services (with several Directorate Generals in charge of the different policy areas involved) in the light of the conclusions of the analysis in the Commission's Staff Working Documents (SWDs). This brings us to an important point, namely, that the role of the fiscal sustainability indicators described in this chapter is very different from that of "fiscal targets" (and the sustainability indicators should therefore not be confused with the monitoring of fiscal rules enshrined in the Stability and Growth Pact). Instead, their role consists in contributing to providing analytical input and support to sound policy advice and recommendations to Member States.

Clearly, the fiscal sustainability indicators provide a very valuable analytical input into the formulation of CSRs in the areas of pensions, healthcare and long-term care policies (see European Commission 2014a). In this context, the S1 and S2 indicators are used to determine whether an important fiscal sustainability challenge exists for the countries under examination. If such challenges are identified, a more specific analysis to establish the nature of the challenge (in terms of the country's pension system, healthcare and long-term care system) is undertaken to make sure that appropriate remedies can be suggested (see European Commission 2014a). Clearly, this type of horizontal screening based on the S1 and S2 indicators does not pre-empt policy, but instead suggests a more detailed analysis of country-specific circumstances, which necessarily lies behind the formulation of policy recommendations on pensions, healthcare and long-term care.

All in all, as explained in detail in this Chapter, in the context of the European Commission's multi-dimensional approach to fiscal sustainability analysis, the S0, S1 and S2 indicators are designed to allow for the early identification of sustainability challenges over the short-, medium- and long-run. Particularly relevant in terms of providing analytical support to policy advice and the formulation of policy recommendations to Member States is the fact that the S1 and S2 indicators can be disaggregated into sub-components that allow identification of the specific areas that challenges stem from (the fiscal stance at the start of the projections and/or the projected dynamics of public spending on pensions, healthcare and long-term care). For the short term, the analysis of the individual variables included in the S0 indicator allows identification of the determinants of the risk signals sent by the overall indicator, which is particularly important to support the devising of appropriate policy responses. These three fiscal sustainability indicators are used to provide analytical input to the SWDs (as well as to the Commission assessment of Member States' Stability and Convergence Programmes) and support the formulation of the Commission's proposed CSRs. In this sense, the role of the sustainability indicators is that of a toolkit to help the reading and understanding of the budgetary situation and the potential fiscal risks for the country under examination, additionally taking into account projected implicit liabilities related to population ageing. Clearly, as explained above, the S1 and S2 indicators provide a very valuable analytical input into the formulation of CSRs in the areas of pensions, healthcare and long-term care policies. It is nonetheless important to stress that the type of horizontal screening based on these two indicators clearly does not pre-empt policy, but instead suggests a more detailed analysis of country-specific circumstances, which necessarily lies behind the formulation of policy recommendations on pensions, healthcare and long-term care.

Notes

1. The European Semester is the EU's integrated annual cycle of economic and budgetary policy coordination, as will be explained better in what follows.
2. In more technical terms, the value of the S2 indicator is derived from the solution of the government inter-temporal budget constraint over the infinite horizon, including its implicit liabilities from ageing, under the assumption that the no-Ponzi game condition is satisfied (i.e. debt and interests on debt are not systematically paid by issuing new debt). The ageing cost projections incorporated in the indicator are the result of joint work between the European Commission and the Member States. The projection results are published every 3 years (see European Commission 2015d, for the latest public release) and are regularly updated with peer-reviewed pension reforms that take place in the meantime.
3. In fact, the initial budgetary position is not defined in exactly the same way for the S1 indicator as for S2, in that S1 includes both a first element that relates to the gap to the debt-stabilising primary balance and a second element that relates to the cost of delaying the fiscal adjustment (the indicator assumes a gradual linear fiscal adjustment taking place over 5 years from the year after the forecast). See European Commission (2015a) for further details.
4. The definition of fiscal stress refers to instances of: (a) very high inflation (above 35%), (b) significant sovereign bond yield spreads (two standard deviations above the mean), (c) public debt default/restructuring/rescheduling and/or (d) a large-scale IMF-supported programme in place.
5. The logic behind the methodology for calculating the thresholds (i.e. the "signals approach") rests on the observation that economies behave in a systematically different way in periods preceding fiscal stress. According to this, time series of the variables used in the analysis (the 28 fiscal and financial competitiveness variables) and the series of fiscal stress episodes recorded in the past are used together to determine an optimal fiscal risk threshold for each of the variables in question, based on its past behaviour ahead of fiscal stress episodes. These optimal thresholds are determined by maximising the "signalling power" of the model, that is, its ability to correctly predict past fiscal stress. By first distinguishing between the two types of errors that can be made in such a prediction (predicting fiscal stress for a variable value beyond the threshold ahead of no fiscal stress episode, type I error, and predicting no fiscal stress for a variable value on the safe side of the threshold ahead of a fiscal stress episode, type II error), the optimal threshold is then determined in a way to minimise the share of missed (in the sense of not signalled) stress episodes plus the share of non-fiscal stress episodes wrongly signalled as upcoming fiscal stress. The thresholds for the S0 indicator and the two fiscal and financial competitiveness sub-indexes are calculated following exactly the same procedure. The thresholds, signalling power and type I and type II errors are reported in Table 10.1 for the S0 indicator, the two sub-indexes and each individual variable.

6. Values of the sustainability indicators are presented here for all the countries for which results were published in the European Commission's 2014 Staff Working Documents.
7. The European Commission's macroeconomic forecasts cover a 2-year horizon. The 2014 SWDs were based on the Commission's spring 2014 forecasts, for which 2015 was the latest forecast year.
8. Based on the medium-term fiscal sustainability indicator S1, countries are classified as: (a) "low risk" if the S1 value is less than zero, (b) "medium risk" if S1 is between 0 and 2.5% and (c) "high risk" if the S1 value is greater than 2.5% (implying a structural fiscal adjustment of more than 0.5% of GDP per year—with the latter representing the benchmark adjustment in the Stability and Growth Pact) (see European Commission 2015a).
9. For the long-term fiscal sustainability indicator S2, the following thresholds are used to assess the scale of the sustainability challenge: (a) if S2 is lower than 2, the country is assigned "low risk"; (ii) if S2 is between 2 and 6, the country is assigned "medium risk"; and (c) if S2 is greater than 6, the country is assigned "high risk" (see European Commission 2015a).
10. The AGS conclusions are discussed and adopted by the Council. The economic priorities based on the AGS are later adopted by the European Council.
11. Starting from the European Semester 2015, the Staff Working Documents and the In-depth Reviews (the reports that follow up the Alert Mechanism Report in the context of the Macroeconomic Imbalances Procedure) were merged into a comprehensive single economic assessment for each Member State, providing the basis for the recommendations to the Member States.
12. Table V in the Annex to the 2014 SWDs, for instance, reported values for all three fiscal sustainability indicators, together with the values of their components.

References

Baldacci, E., Petrova, I., Belhocine, N., Dobrescu, G., & Mazraani, S. (2011). *Assessing Fiscal Stress* (IMF Working Paper, 11/100). Washington, DC: International Monetary Fund. Retrieved from https://www.imf.org/external/pubs/cat/longres.aspx?sk=24822.0

Berti, K., Salto, M., & Lequien, M. (2012). *An Early-Detection Index of Fiscal Stress for EU Countries* (European Economy Economic Paper, 475). Brussels: European Commission. Retrieved from http://ec.europa.eu/economy_finance/publications/economic_paper/2012/pdf/ecp475_en.pdf

Council of the European Union. (2013). *Council Conclusions on the Sustainability of Public Finances in the EU* (3220th Economic and Financial Affairs Council Meeting, 12 February 2013). Brussels: Council of the European Union. Retrieved from http://www.consilium.europa.eu/uedocs/cms_data/docs/pressdata/en/ecofin/135434.pdf

European Commission. (2006). *The Long-term Sustainability of Public Finances in the European Union* (European Economy No. 4/2006). Brussels: Publications Office of the European Union. Retrieved from http://ec.europa.eu/economy_finance/publications/publication7903_en.pdf

European Commission. (2009). *Sustainability Report 2009* (European Economy No. 9/2009). Brussels: Publications Office of the European Union. Retrieved from http://ec.europa.eu/economy_finance/publications/publication15998_en.pdf

European Commission. (2012). *Fiscal Sustainability Report 2012* (European Economy No. 8/2012). Brussels: Publications Office of the European Union. Retrieved from http://ec.europa.eu/economy_finance/publications/european_economy/2012/pdf/ee-2012-8_en.pdf

European Commission. (2014a). *Identifying Fiscal Sustainability Challenges in the Areas of Pension, Health Care and Long-term Care Policies* (European Economy Occasional Paper No. 201/October 2014). Brussels: European Commission. Retrieved from http://ec.europa.eu/economy_finance/publications/occasional_paper/2014/pdf/ocp201_en.pdf

European Commission. (2014b). *Annual Growth Survey 2015* (COM(2014) 902 Final, 28 November 2014). Brussels: European Commission. Retrieved from http://ec.europa.eu/europe2020/pdf/2015/ags2015_en.pdf

European Commission. (2014c). *Staff Working Documents Produced in the Context of the European Semester 2014. Country Specific Recommendations*. Brussels: European Commission. Retrieved from http://ec.europa.eu/europe2020/making-it-happen/country-specific-recommendations/2014/index_it.htm

European Commission. (2015a). *Fiscal Sustainability Report 2015* (European Economy Institutional Paper No. 018/January 2016). Brussels: Publications Office of the European Union. Retrieved from http://ec.europa.eu/economy_finance/publications/eeip/pdf/ip018_en.pdf

European Commission. (2015b). *Alert Mechanism Report 2016* (COM(2015) 691 Final, 26 November 2015). Brussels: European Commission. Retrieved from http://ec.europa.eu/europe2020/pdf/2016/ags2016_alert_mechanism_report.pdf

European Commission. (2015c). *Annual Growth Survey 2016* (COM(2015) 690 Final, 26 November 2015). Brussels: European Commission. Retrieved from http://ec.europa.eu/europe2020/pdf/2016/ags2016_annual_growth_survey.pdf

European Commission. (2015d). *The 2015 Ageing Report: Economic and Budgetary Projections for the 28 EU Member States (2013–2060)* (European Economy No. 3/2015). Brussels: Publications Office of the European Union. Retrieved from http://europa.eu/epc/pdf/ageing_report_2015_en.pdf

Kaminsky, G. L., & Reinhart, C. M. (1999). The Twin Crises: The Causes of Banking and Balance-of-Payments Problems. *American Economic Review, 89*(3), 473–500.

Kaminsky, G., Lizondo, S., & Reinhart, C. M. (1998). Leading Indicators of Currency Crises. *IMF Staff Papers*, *45*(1). Retrieved from http://home.gwu.edu/~graciela/HOME-PAGE/RESEARCH-WORK/WORKING-PAPERS/leading-indicators.pdf

11

Measuring the Rule of Law: The World Justice Project's Rule of Law Index

Alyssa Dougherty, Amy Gryskiewicz, and Alejandro Ponce

Introduction

The United Nations (UN) has asserted that fostering respect for the rule of law around the world is not only 'fundamental to achieving a durable peace in the aftermath of conflict', but that this respect is also integral to the 'effective protection of human rights, and to sustained economic progress and development' (Rule of Law 2016). This statement exemplifies a general consensus within the international community that strengthening the rule of law should be a major goal of governments, donors, businesses, and civil society organisations around the world. Indeed, most scholars and policymakers in recent years have viewed rule of law not only as an instrument for achieving economic development goals but also as an end in and of itself (Ginsburg 2011). This is further exemplified by Goal 16 of the United Nation's

This chapter was prepared for the "Workshop on Global Governance by Indicators – Measuring Governance and Stateness", 13–14 November 2014, in Florence, Italy. It builds on the work developed by Mark Agrast, Juan Botero, and Alejandro Ponce. (For more information about the methodology, see Botero and Ponce 2011; and WJP Rule of Law Index 2016). The views expressed here are those of the authors and do not necessarily represent the views of the WJP, its Board of Directors, or its honorary chairs.

A. Dougherty (✉)
United States Senate, Former World Justice Project, Washington, DC, USA

A. Gryskiewicz • A. Ponce
World Justice Project, Washington, DC, USA

D.V. Malito et al. (eds.), *The Palgrave Handbook of Indicators in Global Governance*,
https://doi.org/10.1007/978-3-319-62707-6_11

Sustainable Development Goals (SDGs), which asserts that rule of law is vital to fostering sustainable development and is a "critical instrument for promoting social cohesion, preventing conflict and ensuring inclusive, safe and peaceful societies" (United Nations Development Programme 2016, p. 5).

However, despite the clear international prominence of rule of law as a conceptual ideal, there still remains little agreement as to what specifically constitutes rule of law, what its current status is, and how to best advance it (Carothers 2006, 2009; Tamanaha 2004; Ghani and Lockhart 2009; Peerenboom et al. 2011). With these difficulties in mind, as an independent multi-disciplinary organisation, the World Justice Project (WJP) set to the task of developing a new framework for conceptualising and assessing the extent to which countries adhere to the rule of law in practice. This framework borrows heavily from theoretical work on the nature of the rule of law and is grounded on the idea that law imposes limits on the exercise of power by government and private interests. In essence, this framework views the rule of law as a two-way relationship between the state and its society. Furthermore, this framework goes beyond establishing general principles and instead explores the ingredients of the rule of law in terms of specific goals, and the ends and outcomes that rule-of-law societies seek to achieve and that policymakers might want to influence. These outcomes are determined by both the formal and informal institutions governing individual and firm behaviour—including the laws under which a society is organised and the incentives these rules create—and by the checks and balances among the power structures within a society that define how government officials conduct themselves and respond to citizens' demands and needs.

This chapter describes the process through which the World Justice Project formulated the *Rule of Law Index*. It discusses the development of the *Index*'s indicators, draws attention to their limitations, and explores the contextual, methodological, and political issues that were taken into consideration prior to its launch. Finally, it highlights how different stakeholders have responded to the *Index*'s indicators and showcases how these indicators have come to impact the conceptual understanding of the rule of law. For further reading on the formation and implementation of the Index, refer to the *WJP Rule of Law Index 2015* report.

Conceptual Framework

Defining the rule of law is an inherently complicated process. The academic literature offers an array of views on what constitutes the rule of law, although none of these views have been universally accepted and adopted internation-

ally. Echoing this point, Carothers (2006) observes that "there is also uncertainty about what the essence of the rule of law actually is" (p. 3). In spite of these observations, however, there is a strong international consensus that the rule of law is a meaningful and important concept. In view of this divergence of understandings, the World Justice Project's team decided—after several iterations, consultations, pilots, and vettings by academic and policy institutions[1]—to weight both theoretical and practical considerations in designing a conceptual framework for the rule of law.

The first consideration was to strike a balance between a "thin" or minimalist conception of the rule of law that focuses on formal procedural rules and a "thick" conception that includes substantive characteristics, such as self-government and various fundamental rights and freedoms. Minimalist or "thin" definitions of the rule of law do not make judgements about the legitimacy or "justness" of laws themselves. Instead, these definitions simply focus on whether rules exist and scrutinise whether these rules are followed by all, including the sovereign. In contrast, substantive definitions take into consideration certain rights that are seen to be fundamental to the rule of law (Carothers 2006, 2009).

When creating a balance between these two approaches, the World Justice Project sought to highlight more than a system of rules or a system of positive law, which usually fails to respect certain core human rights guaranteed under international law. While the divide between these two conceptual approaches to defining the rule of law may reflect deeply rooted cultural differences, a common theme throughout most of these conceptions is that the law imposes limits on the exercise of power by government and private interests (Tamanaha 2004). At its most basic, rule of law refers to a "system in which law is able to impose meaningful restraints on the state and individual members of the ruling elite" (Peerenboom 2002, p. 2). This is captured, for instance, in the definition proposed by the United Nations:

> The rule of law…refers to a principle of governance in which all persons, institutions and entities, public and private, including the State itself, are accountable to laws that are publicly promulgated, equally enforced and independently adjudicated, and which are consistent with international human rights norms and standards. (United Nations Security Council 2004, p. 4)

In a similar vein, the World Justice Project defines the rule of law as:

> a rules-based system in which the following four universal principles are upheld: (1) the government and its officials and agents as well as individuals and private entities are accountable under the law; (2) the laws are clear, publicised, stable

> and just; are applied evenly; and protect fundamental rights, including the security of persons and property and certain core human rights; (3) the process by which the laws are enacted, administered, and enforced is accessible, fair and efficient; and (4) justice is delivered timely by competent, ethical and independent representatives and neutrals who are of sufficient number, have adequate resources, and reflect the makeup of the communities they serve. (Agrast et al. 2011, p. 1)

In the end, the World Justice Project developed a framework that encompassed these essential elements and that emphasised a balance between the thick and thin conceptions of the rule of law. For example, one essential element of the rule of law is the extent to which a country protects fundamental human rights. However, given the impossibility of assessing adherence to the full panoply of civil, political, economic, social, cultural, and environmental rights, the Index addresses a more modest menu of rights—primarily civil and political—that are firmly established under international law and bear the most immediate relationship to rule-of-law concerns.

A second consideration was to create a framework that revolved around the central tenets of both the western and non-western traditions in order to respect diverse interpretations of the rule of law. In spite of the inherent difficulty in comparing different societies, many of the ideas underlying most conceptions of the rule of law are not solely—or even originally—Western, but instead they can be found in an array of legal traditions, institutional architectures, and cultures. For example, the concept that political power must be exercised in accordance with law rather than in an arbitrary or self-interested manner was laid down in the Codex Hammurabi (King 2005). As the *Index*'s success depended on its wide acceptance, this consideration was not merely academic but also practical. The *Index* framework was therefore structured to accommodate countries with vastly different social, cultural, economic, and political systems. For example, the *Index* addresses the extent to which a country provides for fair participation in the making of its laws—certainly an essential attribute of self-government—but it does not address the further question of whether the laws are enacted by democratically elected representatives, as this is a highly contested political issue in many countries.

A third consideration was to focus on the ends of the rule of law rather than its means, as countries can obtain successful rule-of-law outcomes through diverse inputs—including a variety of institutional, legal, and political frameworks. Comparing institutions is not meaningful unless there is an evaluation of their merits or failures across a range of assessment criteria removed from contextual factors. However, the links between inputs and outputs are quite

complex, making it difficult to establish with certainty all the root causes of the multiple weaknesses in the rule of law. Because of this, the Index focuses on rule-of-law outcomes rather than the means by which these results can be obtained—meaning that any specific laws, judiciaries, or law enforcement agencies in place are not acknowledged. Admittedly, this ends-based approach is less actionable, but it provides practitioners with performance information about the outcomes they ultimately want to influence through reform (Kleinfeld 2005, 2012). A corollary of this decision is that the *Index* does not look at the laws as they are written (de jure), but at the consequences arising from the manner in which they are implemented and enforced in practice (de facto). The law becomes meaningful only when it is put into action.

A fourth consideration was to provide a multi-dimensional and comprehensive picture of each country, which prompted the development of an array of indicators that could fully capture the rule of law in practice. This approach separated the *Index* from other indices which covered particular aspects of the rule of law, including the World Bank's *Worldwide Governance Indicators* and *Doing Business Indicators*; Transparency International's *Corruption Perceptions Index*; Freedom House's *Freedom in the World Report*; the American Bar Association Rule of Law Initiative's (ROLI) judicial and other institutional indices; the *Ibrahim Index of African Governance*; and the *Bertelsmann Transformation Index*, which provide valuable information on particular aspects of the rule of law, chiefly as they measure such matters as governance, transparency, investment climate, corruption, and human rights, but never yield a full portrait of rule-of-law compliance.

A final consideration was to place people at the core of the *Index* by looking at a nation's degree of adherence to the rule of law from the perspective of ordinary individuals who are directly affected by it in their societies. With these different considerations in mind, the World Justice Project spent 2 years developing the conceptual framework for the *Index*, which ultimately resulted in the creation of 47 indicators (summarised in Table 11.1 below) organised around nine themes or factors that summarise the outcomes that rule-of-law societies seek to achieve.

The theoretical framework linking these nine themes together is based on two main ideas pertaining to the relationship between the state and the governed, namely, (1) that the law imposes limits on the exercise of power by the state and its agents, as well as individuals and private entities and (2) that the state limits the actions of members of society and fulfils its basic duties towards its population so that the public interest is served, including protecting the people from violence and providing members of society with effective mechanisms to settle disputes and redress grievances. This framework assumes

Table 11.1 The 9 factors and 47 indicators of the WJP Rule of Law Index

Theme	Indicators
1. Constraints on Government Powers	Government powers are effectively limited by the legislature
	Government powers are effectively limited by the judiciary
	Government powers are effectively limited by independent auditing and review
	Government officials are sanctioned for misconduct
	Government powers are subject to non-governmental checks
	Transition of power is subject to the law
2. Absence of Corruption	Government officials in the executive branch do not use public office for private gain
	Government officials in the judicial branch do not use public office for private gain
	Government officials in the police and the military do not use public office for private gain
	Government officials in the legislative branch do not use public office for private gain
3. Open Government	Publicised laws and government data
	Right to information
	Civic participation
	Complaint mechanisms
4. Fundamental Rights	Equal treatment and absence of discrimination
	The right to life and security of the person is effectively guaranteed
	Due process of law and rights of the accused
	Freedom of opinion and expression is effectively guaranteed
	Freedom of belief and religion is effectively guaranteed
	Freedom from arbitrary interference with privacy is effectively guaranteed
	Freedom of assembly and association is effectively guaranteed
	Fundamental labour rights are effectively guaranteed
5. Order and Security	Crime is effectively controlled
	Civil conflict is effectively limited
	People do not resort to violence to redress personal grievances

(*continued*)

Table 11.1 (continued)

Theme	Indicators
6. Regulatory Enforcement	Government regulations are effectively enforced
	Government regulations are applied and enforced without improper influence
	Administrative proceedings are conducted without unreasonable delay
	Due process is respected in administrative proceedings
	The government does not expropriate without lawful process and adequate compensation
7. Civil Justice	People can access and afford civil justice
	Civil justice is free from discrimination
	Civil justice is free from corruption
	Civil justice is free from improper government influence
	Civil justice is not subject to unreasonable delay
	Civil justice is effectively enforced
	ADR is accessible, impartial, and effective
8. Criminal Justice	The criminal investigation system is effective
	The criminal adjudication system is timely and effective
	The correction system is effective in reducing criminal behaviour
	The criminal system is impartial
	The criminal system is free from corruption
	The criminal system is free from improper government influence
	Due process of law and rights of the accused
9. Informal Justice	Informal justice is timely and effective
	Informal justice is impartial and free from improper influence
	Informal justice respects and protects fundamental rights

Source: Reprinted from Agrast, M., Botero, J., and Ponce, A., WJP Rule of Law Index (2011)

very little about the functions of the state, and when it does so it incorporates functions that are recognised by all societies, such as the provision of justice or the guarantee of order and security. The framework is grounded in classic social contract theories. It incorporates elements of the philosophies of Thomas Hobbes and John Locke, who argued that the fundamental role of government is to provide security, peace, and defence in a civil society. It also includes elements from James Madison, who established that every form of government has to have a system to ensure that no one in the government has so much power that they can act above the law. Finally, it also incorporates

elements of Max Weber's theory of bureaucracy and state administration by highlighting the efficiency, objectivity, and effectiveness of state officials in carrying out their duties.

The first four factors emphasise the idea that state actors are bound by law and should be kept in check to prevent abuse of power. The first of these comprises the institutional means by which the powers of the government and its officials are limited and by which they are held accountable under the law. The second measures the absence of corruption.[2] The third regards transparency, openness, and participation in government, and the fourth factor measures protection of fundamental human rights. Factors 5–8 encapsulate the notion that the state has a duty to protect the people from violence (factor 5), regulate the actions of members of society (factor 6), and run courts that resolve disputes and enforce contracts (factor 7) and that redress grievances and bring action against individuals for offences against society (factor 8). Finally, factor 9 concerns the role played in many countries by traditional, or "informal", systems of law—including traditional, tribal, and religious courts as well as community-based systems—in resolving disputes. These systems often play an important role in cultures in which formal legal institutions fail to provide effective remedies for large segments of the population, or when formal institutions are perceived as foreign, corrupt, and ineffective (Pimentel 2010; Irfan 2009). These factors are further disaggregated into 47 sub-factors, which together provide a comprehensive picture of the rule of law.[3]

Measuring the Rule of Law

Following the development of the *Index*'s conceptual framework described above, the World Justice Project took steps to produce each of the 47 indicators. In order to produce valid measures, the Project considered three methodological approaches and carried out a series of pilots to test the soundness of the various data-collection procedures.

Approaches

The first approach aimed at measuring each of the 47 sub-factors using existing cross-country data sources on institutions, governance, corruption, human rights, transparency, and justice.[4] The second approach attempted to populate some of the indicators with in-country data produced by NGOs, statistical agencies, and other institutions. This pilot exercise was conducted by the Vera

Institute of Justice for the WJP and was tested in Chile, India, Nigeria, and the United States. The exercise focused on criminal justice indicators and drew on a range of data sources, including information from the police, courts, and prisons, as well as NGO reports and legislation. The third approach combined different data-collection methods and sources of information, including a standardised general population poll, expert surveys, and analyses of cross-country data from existing third-party sources. This methodology was developed by the World Justice Project team and tested in Argentina, Australia, Colombia, Spain, Sweden, and the United States in 2008.

While each of these methodologies has strengths and weaknesses, some may be better suited than others to the particular needs of various audiences. The first approach, for example, is cost-effective and can yield scores for a large number of countries relatively fast, but it has many limitations in terms of thematic and country coverage. The second approach is grounded on hard data, but it raises methodological concerns about data comparability. This approach also faces significant practical difficulties if there is to be a rapid and standardised expansion of the number of countries that can be covered. The third approach is anchored on in-country data sources—including a poll of the general population—and generates indicators that are much more comparable across countries, but it demands excessive resources and the participation of a large number of highly qualified local individuals.

The decision on how to weight considerations about country coverage, thematic coverage, data-collection methods, frequency, cost, replicability, and cultural competency depends on the intended audiences and goals of the *Index*. From a conceptual point of view, there is a tension among the needs of various users of information. While all of them seek timely and accurate information, they each have different goals and emphases. For instance, the business community seeks simplicity and flexibility to enable rapid decision-making, while the legal community looks for conceptual precision and detail. Similarly, within the academic community, while development economists expect comparability and standardisation, sociologists and anthropologists focus on understanding local realities and cultural competency. In choosing a methodological approach, the World Justice Project team weighed all these considerations against other practical ones, such as building a measurement tool that could facilitate communication of the status of the rule of law to civil society and policymakers, promoting accountability, encouraging dialogue, and placing country performance and progress at the centre of policy discussions.

Ultimately, the World Justice Project opted for the third approach and decided to poll a representative sample of the general public in each country

and complement these data with the perceptions of local practitioners in four fields: civil and commercial law, criminal justice, labour law, and public health. The idea was to obtain first-hand information on the experiences and perceptions of those affected by the level of adherence to the rule of law in a country. Collecting primary data has important advantages, as it allows for an increased precision in the multiple concepts measured and for a neat transition from concepts to variables to indicators, avoiding situations where the availability of data actually drives the concepts measured. Collecting information from different data sources, on the other hand, allows for different views and an expansion of knowledge. Experts and lay people are knowledgeable about different rule-of-law situations. Experts, for instance, have specialised knowledge of certain processes, actors, institutions, and circumstances, while the general public possess first-hand experience of real-life situations resulting from the level of adherence to the rule of law in their community (e.g., having to pay a bribe to access health services, facing police discrimination, or confronting barriers to have access to justice). Using two data sources also serves to validate the findings by providing different perspectives on the same issue. Not only does this approach take into account different perspectives on the rule of law, but it also helps reduce possible biases introduced by any one particular method of collecting data (such as social desirability bias). Therefore, the *Index* anchored expert opinion to rigorous polling of the general public to ensure that the findings reflected the conditions experienced by the population, including marginalised sectors of society.

Data Sources and Score Computation

The first data source utilised by the World Justice Project is a general population poll (GPP). The poll was designed to provide information on the experiences and perceptions of ordinary people concerning their dealings with the government, the police, and the courts; the openness and accountability of the state; the extent of corruption; and the magnitude of common crimes to which the general public are exposed. These data are collected with the assistance of leading local polling companies around the world. The general population polls are carried out on probability samples of 1000 adult respondents drawn from the three largest cities in each country using a multi-stage random procedure. The sampling framework varies by country and is chosen in consultation with the polling company to produce valid representative samples. Depending on the country, the polls are conducted using computer-assisted

telephone interviewing (CATI), face-to-face[5] or online methodologies based on panels produced by non-internet methods to ensure random selection. Before the full implementation of the GPP, the polling company runs a pilot in each country to test the questionnaire and, during fieldwork, follows various procedures to ensure quality control.

The qualified respondent questionnaires (QRQs) complement the polling data with assessments by in-country professionals with expertise in civil law, criminal law, labour law, and public health. These questionnaires gather input from practitioners who frequently interact with state institutions. The questionnaires contain closed-ended perception questions and several hypothetical scenarios based on highly detailed factual assumptions aimed at ensuring comparability across countries. In each country the QRQ surveys are answered by an average of 25 law professors and practising attorneys with significant practical experience in at least one of the four areas mentioned above. These respondents are selected through a two-stage procedure using directories and referrals.

The questions for both the GPP and the QRQ are formulated in closed-ended format to facilitate both data-collection and cross-country comparisons. The QRQ surveys are administered on a yearly basis in each country surveyed, and the GPPs are carried out every 2 years. Together, these two data sources contain more than 500 variables that capture a variety of situations and perspectives.[6] These variables examine practical and concrete situations that reflect the rule-of-law situation in the country and that relate to the factors and sub-factors in the *Index*, as well as perceptions that are relevant to evaluating the rule of law or the performance of state institutions. A total of 35 countries were polled in 2010, 66 in 2011, 97 in 2013, 99 in 2014, and 102 in 2015.[7] For these reports, the countries were selected sequentially to ensure diversity and representation of all the regions, income levels, population sizes, and legal traditions of the world. Methodological challenges were also taken into consideration. The *WJP Rule of Law Index 2015* report was based on data collected by the WJP regarding the experiences and perceptions of more than 100,000 people and 2,500 practitioners.

Once collected, the data are cleaned and processed by mapping each question to its corresponding indicator, scaling the questionnaire items so that all values fall between 0 (least rule of law) and 1 (most rule of law), and aggregating individual responses at the country level using simple averages. The resulting scores are normalised and aggregated from the variable level all the way up to the factor level to produce the final country scores and rankings (see Methodology 2017). In addition to the annual scores, the World Justice

Project conducts statistical tests every year to evaluate the significance of the annual difference in country scores.

Validity in Data Collection and in Building Indicators

One methodological issue in collecting primary data is ensuring that the measures are valid. Some of the possible threats to validity in cross-country studies include comparability of questions and scales, question content and consistency, expert bias,[8] cross-country comparability, and sensitive questions that may be perceived as threatening by government officials or by respondents.[9] The World Justice Project paid attention to developing questions and scales that could easily be understood by respondents and could work in all countries.[10] At the same time, the World Justice Project's team annually evaluates any need for alternative sampling methods or sample sizes in extreme situations, and, if necessary, identifies interview methods that could encourage survey participants to respond accurately.

Another methodological issue is related to the validity of the indicators. For this reason, the *Index* relies on a large number of questions to approximate each one of the concepts and incorporates two different data sources.[11] In addition, the indicators are validated and cross-checked against qualitative and quantitative third-party sources to provide an additional layer of analysis, and to identify possible mistakes or inconsistencies within the data.[12] Finally, the World Justice Project has conducted sensitivity analyses together with the EU Joint Research Centre to test how variation in the assumptions (including missing data, weighting, normalisation, and aggregation) and the samples (to account for sampling error) can alter the *Index* scores and rankings. The results over the years show that country classifications across the nine factors are robust to changes in the modelling assumptions (90% of the countries shift less than ±1 position).

Communicating the Results

Given the aforementioned considerations, a trade-off quickly emerged between producing a single country score and generating disaggregated scores. While a single score for a particular country was attractive to the media and could be easily publicised, it masked the richness of the data collection, limited in-depth analysis, and discounted the complexity of everyday rule-of-law situations. The *Index* emphasised disaggregated data, which helped identify

both the strengths and weaknesses of a country surveyed and could enhance debate and thereby produce a clearer image of each country's rule-of-law landscape. By producing a comprehensive and multi-dimensional picture of the extent to which each country adhered to the rule of law in practice which was benchmarked against comparable countries (in the same region or in the same income group), the World Justice Project avoided shaming countries with lower scores, making blatant comparisons between a particularly developed country against a developing one, or penalising them based on their limited economic development, the complexity of their cultural differences, or their potentially weak legal architectures. The World Justice Project's focus has always been on being honest brokers of information and engaging with dialogue and data—rather than an agenda—and has maintained the organisation's independent and apolitical stance. This approach has been useful to engage various stakeholders in evidence-based dialogue, as we discuss below.

Stakeholder Responses

Since the launch of the *Index*, the World Justice Project has contacted media outlets and civil society organisations in every country indexed in order to share the results with as many citizens as possible. As of September 2016, the *Rule of Law Index* has been cited by more than 2,200 media outlets in nearly 125 countries,[13] and referenced by chief justices, business leaders, and public officials around the world. The Index has been hailed by academics (Bedner 2010, e.g., considers the *WJP Rule of Law Index* to be "the most sophisticated rule of law indicators developed so far") and cited as a means of measuring and defining the rule of law, thereby contributing to the goal of developing a universal definition. It has also been incorporated into global conversations (such as the discussions on the UN Post 2015 Millennium Development Goals), included in reports produced by the OECD (2015) and the European Union (EU Justice Scoreboard 2013), and utilised as a data source in other indices, including Transparency International's *Corruption Perception Index* (Corruption Perception Index 2014 in details 2014) and the World Bank's *Worldwide Governance Indicators* (2017).

Change comes both bottom up and top down. While it is difficult to assess the long-term outcomes of steady trust-building and educational efforts, we have seen multiple examples of our message hitting home, with various governments and individuals pushing for reforms over the past year. In Pakistan in late 2015, parliamentarians formed a multi-party caucus focused specifically on the rule of law, citing the country's low ranking in

the *Index* as evidence that rule-of-law issues must be further addressed. "Parliamentarians on Thursday formed a multi-party caucus on rule of law, said a press release. Under the convenership of Senator Hasil Khan Bizenjo, the caucus comprises parliamentarians from both the Senate and the National Assembly. Senator Bizenjo stated that it was alarming that the sixth most populous country in the world was ranked a dismal 98th out of a total of 102 countries on the World Justice Project Report on *Rule of Law Index 2015*" (Rule of law: MPs' caucus formed 2015). Similarly, in 2015 members of South Africa's Parliament pointed to *Index* findings to highlight issues of corruption and gaps in the country's capacity to impose sanctions for official government misconduct (see South African MP Mkhuleko Hlengwa 2015). The vice president of Malaysia's People's Justice Party (PKR), Shamsul Iskandar Mohd Akin, referenced Malaysia's ranking in the Index and specifically pointed to the decrease in the country's global ranking when he called on the government to commit to upholding the rule of law as part of its national development initiatives (see Free Malaysia Today Reporters 2015).

The *Rule of Law Index* has also been received with interest by heads of state, with its positioning being largely dependent on a country's political, economic, and social situations. Some governments and political leaders have publicly endorsed the *Index* to highlight their country's impressive rankings. Others have used the Index to call for rule-of-law reform, while a few others have attacked the *Index* as they disagree with its findings, often employing anecdotal evidence to strengthen their arguments. A few examples illustrate these points: (a) Georgian Prime Minister Irakli Garibashvili praised Georgia's first-place ranking among the Eastern Europe and Central Asian countries in the *Open Government Index* via his official Twitter account (Garibashvili 2015); (b) Colombian President Juan Manuel Santos used the *Index* findings to underscore the need to reform the Colombian criminal investigation system (El Espectador 2010), a reform that had been delayed 15 years; (c) Vice President of the Philippines Jejomar C. Binay cited the *Index* scores for his country when advocating stronger protection of intellectual property rights (Binay 2011)[14]; (d) Chief Justice of Pakistan Iftikhar Muhammad Chaudhry recognised the need for an objective mechanism to measure Pakistan's adherence to the rule of law and declared that the *WJP Rule of Law Index* should be utilised in this regard (The News 2012); (e) Chief Justice of Canada Beverley McLachlin used the *Index* findings to emphasise the need to address access to legal counsel issues in civil disputes;[15] and, finally, (f) Bangladesh's Law Minister Anisul Huq claimed that, in his view, the WJP's findings were not correct (The Daily Star 2014).

In addition, many other governments have engaged with the World Justice Project, inquiring about the *Index*'s data and its capacity to serve as a monitoring tool.

Finally, the *Index* has also been utilised directly or indirectly (through the use of data sources that incorporate the *Index* findings) as a reliable source of data by funding organisations such as the Millennium Challenge Corporation, and by rating agencies to evaluate the performance of countries (see Moody's 2016). In spite of all this evidence, additional work is needed to evaluate how, why, and the extent to which these types of indicators are actually used to drive discussions and shape policy reforms (AidData 2015).

Conclusion

Rule-of-law development requires clarity about the fundamental features of the rule of law, and an adequate basis for its evaluation and measurement. The *WJP Rule of Law Index* seeks to embody these outcomes within a simple and coherent framework that is broadly accepted, and to approximate by means of performance indicators the extent to which countries need to achieve these outcomes. By capturing the experiences and perceptions of both ordinary citizens and in-country professionals concerning the performance of the state and its agents, and the actual operation of the legal framework in their country, the *WJP Rule of Law Index* aims to give voice to a wide range of individual experiences and concerns so as to allow for more informed assessments and responses. In addition, by operationalising the rule of law into specific concepts and concrete questions, the World Justice Project seeks to create opportunities for fair country comparisons and encourage dialogue to help advance the rule of law worldwide.

The *Index*'s findings, however, should be interpreted in the light of various inherent limitations, including differences in value structures, goals, and legal architectures across countries; measurement error in questions; sensitive questions in certain countries; and urban sampling. More importantly, the *Index* provides a simplified picture and should therefore not be considered a sole means of measuring a country's adherence to effective rule of law. Instead, it should be used with other data—such as hard in-country data, results from quantitative and qualitative research, or journalism—to derive accurate and meaningful conclusions and further engagement with civil society. Policymaking in the area of rule of law requires careful consideration of all the relevant dimensions, and pre-existing data—which may vary from country to country—and a combination of sources, instruments, and

methods are essential in attempting to capture a broad view of the rule of law in a nation.

Notes

1. The World Justice Project organised a series of regional meetings that were a particularly rich source of feedback and advice. The initial draft, Beta Test Version 1.0, was presented in February 2007 in Washington, D.C. Subsequent beta versions were presented at international multi-disciplinary outreach meetings in the Czech Republic, Singapore, Argentina, and Ghana between July 2007 and January 2008, bringing together some 200 individuals from more than 15 disciplines and 61 nations. The World Justice Project also organised academic seminars at Stanford University and Yale University to discuss the conceptual framework and the methodology. The participants at these meetings were invited to scrutinise the structure of the Index, rule of law definitions, and applicable international standards, cultural competencies, the applicability of the Index to diverse legal systems, the degree to which the Index should attempt to assess informal systems of law, the design of rule-of-law indicators and proxies, and methodological issues related to the measurement, testing, and analysis of the results. The participants at these meetings and seminars provided a wide range of comments and criticisms that were extraordinarily valuable in helping to ensure that the *Index* is applicable to societies with diverse social, political, and legal systems, to correct for cultural bias, and to anticipate and address methodological concerns.
2. Governments regulate markets and tax citizens and firms to provide public goods. This opens the possibility of corruption, standardly defined as the abuse of public office for private gain, and selective enforcement.
3. A careful examination of the nine factors shows that there is a partial overlap among some sub-factors. This is because various rule-of-law dimensions partially overlap in practice. For example, a free press is both a manifestation of a fundamental right in action and a non-governmental check on the government's powers.
4. Some of the most relevant cross-country sources considered include the Bertelsmann Foundation (Transformation Index); Brown University (Center for Public Policy: Global E-Government Index); CEELI (CEDAW, Convention to Eliminate all forms of Discrimination against Women Assessment); CEELI (JRI, Judicial Reform Index); CEELI (LPRI, Legal Profession Reform); CEELI (Prosecutorial Reform Index); CEELI (ICCPR Legal Implementation Index, International Covenant on Civil and Political Rights); Center for Systemic Peace (Polity IV Project); EBRD (Sector Specific Assessment of Law and Practices); EBRD (Transition Report); the European Bank for Reconstruction

and Development, EBRD (Country Law Assessments); Freedom House (Countries at the Crossroads); Freedom House (Freedom in the World); Freedom House (Nations in Transit); Global Insight (Global Risk Service); IJET (Country Security Ratings); Institute for Management Development (World Competitiveness Yearbook); International Budget Project (Open Budget Initiative); International Research and Exchange Board (Media Sustainability Index); New Tools in Comparative Political Economy: The Database of Political Institutions; Political and Economic Risk Consultancy (Asian Intelligence: Corruption Report); Political Risk Service (International Country Risk Guide); Public Financial Management; Reporters Without Borders (Press Freedom Index); Russell's EMPulse, Investors' Perceptions of the Pulse of Emerging Markets; The Global Integrity Report; Transparency International (Corruption Barometer); USAID (NGO Sustainability Index for Central and Eastern Europe and Eurasia); World Bank (Country Policy and Institutional Assessments); World Bank (Doing Business); World Bank (DPI, Database of Political Institutions 2006); Afro-barometer; CIMA (Barómetro Iberoamericano de Gobernabilidad); Gallup World Poll; Global Insight (Economic and Financial Data); Heritage Foundation (Index of Economic Freedom); Latinobarometro; Transparency International (Bribe Payers Index); US State Department (Trafficking in Persons Report); Vanderbilt University (LAPOP, The Americas Barometer); World Bank (Enterprise Surveys); World Economic Forum (The Global Competitiveness Report); Cingranelli-Richards (CIRI, Human Rights Dataset); Mo Ibrahim Foundation (Ibrahim Index of African Governance); Organisation for Economic Co-operation and Development (OECD, African Economic Outlook); Political Terror Scale; Transparency International (CPI, Corruption Perceptions Index); World Bank (Worldwide Governance Indicators); African Development Bank (Country Policy and Institutional Assessments); Asian Development Bank (Country Policy and Institutional Assessments); Amnesty International Report; Economist Intelligence Unit (Country Risk Service and Country Forecasts); Human Rights First (Annual Report); Human Rights Watch (Country Reports); Open Society Institute and EU Monitoring and Advocacy Program (EU Accession Reports); United Nations (UN, Universal Human Rights Index); and the United States Department of State (Human Rights Practices Annual Report).

5. In the case of CATI and face-to-face methodologies, the polling company makes three contact attempts before substituting a respondent.
6. In addition to these data sources, the Index incorporates third-party data to measure structural rule-of-law situations that may not be captured through general population polls or expert opinions. These variables include (a) the number of events and (b) the number of deaths resulting from high-casualty terrorist bombings (see Center for Systemic Peace, Major Episodes of Political Violence, 1946–2015); (c) the number of battle-related deaths; (d) the num-

ber of casualties resulting from one-sided violence [Source: Uppsala Conflict Data Program]; and (d) coup d'état events [Coded from the Center for Systemic Peace].

7. The 35 countries covered in the WJP Rule of Law Index 2010 report were Albania, Argentina, Australia, Austria, Bolivia, Bulgaria, Canada, Colombia, Croatia, Dominican Republic, El Salvador, France, Ghana, India, Indonesia, Japan, Jordan, Kenya, Republic of Korea, Liberia, Mexico, Morocco, Netherlands, Nigeria, Pakistan, Peru, Philippines, Poland, Singapore, South Africa, Spain, Sweden, Thailand, Turkey, and the United States. The following year, the 2011 report included, in addition to the previous countries, Bangladesh, Belgium, Brazil, Cambodia, Cameroon, Chile, China, Czech Republic, Estonia, Ethiopia, Germany, Guatemala, Hong Kong SAR, China, Iran, Italy, Jamaica, Kazakhstan, Kyrgyzstan, Lebanon, Malaysia, New Zealand, Norway, Romania, Russia, Senegal, Uganda, Ukraine, United Arab Emirates, the United Kingdom, Venezuela, and Vietnam. For the 2012 report, in addition to the 66 previous countries, the Index report provided scores for Belarus, Bosnia and Herzegovina, Botswana, Burkina Faso, Cote d'Ivoire, Denmark, Ecuador, Egypt, Finland, Georgia, Greece, Hungary, Macedonia, FYR, Madagascar, Malawi, Moldova, Mongolia, Nepal, Nicaragua, Panama, Portugal, Serbia, Sierra Leone, Slovenia, Sri Lanka, Tanzania, Tunisia, Uruguay, Uzbekistan, Zambia, and Zimbabwe. In addition to these countries, the 2014 report included scores for Afghanistan and Myanmar. Finally, the 2015 report included scores for all the countries previously mentioned plus Belize, Costa Rica, and Honduras for a total of 102 countries.
8. Experts may not be exposed to certain problems that the general public experience or may be biased against certain policies or forms of government.
9. The government officials of certain countries may censor or condition the administration of questions because they are perceived as challenges to the regime. In other cases, respondents may be unwilling to truthfully answer some questions either because of a perceived risk of sanctions or because of concerns about self-image when the documented behaviour does not conform to social norms (social desirability bias). This can lead to under-reporting of sensitive topics, thus making the data less valid.
10. The World Justice Project conducted a thorough review of about 30 surveys touching on concepts related to the Index. Some survey questions were directly incorporated into the GPP questionnaire; some survey questions were included in a modified form, and others were simply used to inform the design of new questions. The questionnaires also include vignettes, which allowed us to standardise the frame of reference for respondents around the world and as many experience questions as possible as perceptions of institutional performance may not be comparable or may not reflect actual experiences with the system.

11. The rule of law is a complex phenomenon, which may be perceived and experienced differently by different persons depending on their environment, background, positioning, professional expertise, attitudes, ideological tendencies, or beliefs concerning certain topics. Relying on different sources provides a richer picture of the rule-of-law situation in a country.
12. In the case of qualitative checks, we gather relevant quotes from studies such as the United States Human Rights Report, Freedom House's Nation in Transit, Amnesty International country reports and Freedom House's Freedom in the World. This information is gathered in an internal document and compared against our estimates. For the quantitative checks, we collect cross-country data from a large number of sources including the Global Integrity Report, WEF Global Competitiveness Report, Freedom House, WB Governance Indicators, Transparency International's Corruption Perception Index, Gallup, the WJP Doing Business Report, and the Cingranelli-Richards Human Rights Database.
13. The number of articles citing or featuring WJP and the *Index* rose from a total of 279 in 2011 to nearly 1000 in 2015. WJP findings were cited in media outlets in 98 countries in 2015, up from 49 countries in 2011.
14. "An effective campaign against piracy and counterfeiting is also seen as further strengthening the rule of law. While the Philippine Development Plan 2011–2016 cites the rule of law as vital to the holistic development of citizens, the World Justice Project Rule of Law Index 2010 sadly ranks the Philippines last or close to the bottom among seven indexed Asian countries…", Vice President of the Philippines Jejomar C. Binay (Binay 2011).
15. "In its 2011 Rule of Law Index, the World Justice Project surveys 66 countries to assess the state of the rule of law in each of these jurisdictions…On access to justice, the index ranks Canada 9th out of 12 wealthy Western European and North American countries. The most problematic areas, according to the index, are access to legal counsel and unreasonable delay in civil justice" Chief Justice of Canada Beverley McLachlin (McLachlin 2011).

References

Agrast, M., Botero, J., & Ponce, A. (2011). *WJP Rule of Law Index 2011*. Washington, DC: The World Justice Project.

AidData. (2015). *Report 2105*. Williamsburg: AidData. Retrieved from http://aiddata.org/reports-and-working-papers

Bedner, A. (2010). An Elementary Approach to the Rule of Law. *Hague Journal on the Rule of Law, 2*(1), 48–74.

Binay, J. C. (2011). *Speech of Vice President Jejomar C. Binay During the 1st Philippine Anti-Counterfeiting and Piracy Summit*. Retrieved from https://www.facebook.com/notes/jojo-binay/speech-of-vice-president-jejomar-c-binay-during-the-1st-philippine-anti-counterf/290429334315238/

Botero, J. C., & Ponce, A. (2011). *Measuring the Rule of Law* (The World Justice Project Working Paper Series). Retrieved from https://ssrn.com/abstract=1966257 or http://dx.doi.org/10.2139/ssrn.1966257

Carothers, T. (Ed.). (2006). *Promoting the Rule of Law Abroad: In Search of Knowledge*. Washington, DC: Carnegie Endowment for International Peace.

Carothers, T. (2009). Rule-of-Law Temptations. In J. J. Heckman, R. L. Nelson, & L. Cabatingan (Eds.), *Global Perspectives on the Rule of Law* (pp. 49–61). New York: Routledge.

Corruption Perception Index. (2014). *Corruption Perception Index 2014: In Detail.* Berlin: Transparency International. Retrieved from http://www.transparency.org/cpi2014/in_detail

El Espectador. (2010, October 15). *Gobierno le pide a la Corte más proactividad para reformar la justicia*. Retrieved from http://www.elespectador.com/noticias/politica/articulo-229897-gobierno-le-pide-corte-mas-proactividad-reformar-justicia

EU Justice Scoreboard. (2013, March 27). *EU Justice Scoreboard: European Commission Broadens the Scope of Its Analysis of Member States' Justice Systems* [Press Release]. Brussel: European Commission. Retrieved from http://europa.eu/rapid/press-release_IP-13-285_en.htm

FMT Reporters. (2015, June 16). People's Justice Party Vice President Shamsul Iskandar Mohd Akin. *Free Malaysia Today*. Retrieved from: http://www.freemalaysiatoday.com/category/nation/2015/06/16/malaysias-low-ranking-in-rule-of-law-index-worrying/

Garibashvili, I. (2015, June 12). Georgia Is Number 1 in Eastrer Europe and Central Asia and 29th in the world. *#ROLindex from @TheWJP* [Twitter Moment]. Retrieved from https://twitter.com/GovernmentGeo/status/609332973662384128

Ghani, A., & Lockhart, C. (2009). *Fixing Failed States: A Framework for Rebuilding a Fractured World*. Oxford: Oxford University Press.

Ginsburg, T. (2011). Pitfalls of Measuring the Rule of Law. *Hague Journal on the Rule of Law, 3*, 269–280.

Hlengwa, H. M. (2015). South Africa: Parliamentary Debate – The Ongoing Scourge of Corruption in SA and Measures to Combat This [Press Release]. *All Africa*. Retrieved from http://allafrica.com/stories/201509090566.html

Irfan, H. (2009). "Honor" and Violence Against Women in Pakistan. In Y. Ghai & J. Cottrell (Eds.), *Marginalized Communities and Access to Justice* (pp. 162–184). New York: Routledge.

King, L. W. (2005). *The Code of Hammurabi: Translated by L. W. King*. New Haven: Yale Law School Lillian Goldman Law Library. The Avalon Project. Documents in Law, History and Diplomacy. Retrieved from http://avalon.law.yale.edu/ancient/hamframe.asp

Kleinfeld, R. (2005). *Competing Definitions of the Rule of Law: Implications for Practitioners* (Carnegie Papers, Rule of Law Series, 55). Washington, DC: Carnegie Endowment for International Peace.

Kleinfeld, R. (2012). *Advancing the Rule of Law Abroad: Next Generation Reform*. Washington, DC: Brookings Institution Press.

McLachlin, B. (2011, August 16). *Beverly McLachlin: A Busy Court, Access to Justice, and Public Confidence*. Retrieved from http://ipolitics.ca/2011/08/16/beverley-mclachlin-address-to-the-council-of-the-canadian-bar-association/

Moody's. (2016). *Moody's: Pakistan's B3 Rating Reflects Strengthening Growth, Progress on Structural Reforms*. New York: Moody's Corporation. Retrieved from https://www.moodys.com/research/Moodys-Pakistans-B3-rating-reflects-strengthening-growth-progress-on-structural--PR_347953

OECD. (2015). *Government at a Glance 2015*. Paris: OECD Publishing. doi: http://dx.doi.org/10.1787/gov_glance-2015-en

Peerenboom, R. (2002). *China's Long March Toward Rule of Law*. New York: Cambridge University Press.

Peerenboom, R., Nollkaemper, A., & Zürn, M. (2011). Conclusion: From Rule of Law Promotion to Rule of Law Dynamics. In A. Nollkaemper, R. Peerenboom, & M. Zürn (Eds.), *Rule of Law Dynamics: In an Era of International and Transnational Governance* (pp. 305–324). Cambridge: Cambridge University Press.

Pimentel, D. (2010). Rule of Law Reform Without Cultural Imperialism? Reinforcing Customary Justice Through Collateral Review in Southern Sudan. *Hague Journal on the Rule of Law, 2*(1), 1–28.

Tamanaha, B. (2004). *On the Rule of Law: History, Politics, Theory*. Cambridge: Cambridge University Press.

The Daily Star. (2014, March 8). Bangladesh Ranks 92nd in Rule of Law Index. *The Daily Star*. Retrieved from http://www.thedailystar.net/bangladesh-ranks-92nd-in-rule-of-law-index-14633

The Express Tribune. (2015, September 18). Rule of Law: MPs' Caucus Formed. *The Express Tribune*. Retrieved from https://tribune.com.pk/story/958795/rule-of-law-mps-caucus-formed

The News. (2012). Pak Money Laundering Laws Should Conform to Global Standards. *The News*. Retrieved from https://www.thenews.com.pk/archive/print/621707-pak-money-laundering-laws-should-conform-to-global-standards

The World Justice Project (WJP). (2016). *World Justice Project Rule of Law Index*. Washington, DC: The World Justice Project. Retrieved from http://worldjusticeproject.org/rule-of-law-index

The World Justice Project (WJP). (2017). *Methodology*. Washington, DC: World Justice Project. Retrieved from http://worldjusticeproject.org/methodology

United Nations Development Programme (UNDP). (2016). *UNDP Support to the Implementation of Sustainable Development Goal 16*. New York: United Nations Development Programme. Retrieved from http://www.undp.org/content/dam/undp/library/Sustainable%20Development/16_peace_Jan15_digital.pdf?download

United Nations Peacekeeping. (2016). *Rule of Law*. Unite Nations Peacekeeping. Retrieved from http://www.un.org/en/peacekeeping/issues/ruleoflaw.shtml

United Nations Security Council. (2004, August 23). *The Rule of Law and Transitional Justice in Conflict and Post Conflict Societies: Report of the Secretary General*. New York: United Nations Security Council. Retrieved from https://www.un.org/ruleoflaw/blog/document/the-rule-of-law-and-transitional-justice-in-conflict-and-post-conflict-societies-report-of-the-secretary-general/

Worldwide Governance Indicators (WGI). (2017). [Website]. Retrieved from http://info.worldbank.org/governance/wgi/index.aspx#doc

12

Measuring the Opposite of Corruption: The Evolution of Governance Indicators at Global Integrity

Hazel Feigenblatt and Johannes Tonn

Introduction

Global Integrity is known for producing data on integrity and anti-corruption in various data sets, notably the Global Integrity Report (GIR) and the Africa Integrity Indicators (AII). This chapter explains why and how Global Integrity came to contribute to the governance and, in particular, to the anti-corruption field, and how our methodology and approach have continued to evolve over the past decade. This is not the first time that we publicly reflect on our approach and methodology (see Heller 2011a; Hudson 2015) and we are thankful for the opportunity to articulate our thinking and to share it with a wider audience. We welcome feedback on our reflections and invite discussion about our assumptions, strategy and the lessons learned.

Three important stages mark the evolution of our approach and the associated methodology over time. The first one involves Global Integrity's origin, which was rooted in a reflection about the state of the governance field and the existing data sets and prevalent methodologies in the early 2000s. The second stage—starting around the year 2011—was marked by a thorough revision and systematic refinement of our methodology and an expansion of

H. Feigenblatt (✉)
University of Maryland, Washington, DC, USA

J. Tonn
Global Integrity, Washington, DC, USA

D.V. Malito et al. (eds.), *The Palgrave Handbook of Indicators in Global Governance*,
https://doi.org/10.1007/978-3-319-62707-6_12

our work into new fields, based on our continued exploration and better understanding of the many possible methodological approaches, trade-offs and ensuing results. Finally, and based on our evolving understanding of how governance reform happens in practice, we are in the midst of a third stage—starting in 2015—which is motivated by considerations about better understanding and supporting the use and usefulness of data at the country level to support domestic actors in driving governance reform.

As others have documented (Malito 2014; Cooley 2015), there has been exponential growth in the production of governance assessments and Global Integrity has been fortunate to play a key role in what the World Bank calls the "second wave" of governance assessments (World Bank 2016a, b). This "second wave" follows a first wave, mainly composed of widely recognised quantitative composite indices such as the Worldwide Governance Indicators and Transparency International's Corruption Perceptions Index. Indeed, and as we explain later in this chapter, part of our motivation for creating the GIR was to counter some of the unresolved challenges innate to the first wave of governance indicators.

Following a brief excursion into the genesis of the GIR (Global Integrity 2006–2013) and its underlying rationale, we explain why we decided to measure the "opposite of corruption" (i.e. integrity) and how we improved our indicator design and research process over time, with our methodology eventually evolving from a traditional expert-based assessment to "fact-based expert analysis".

We have been, and continue to be, acutely aware of the risks and responsibilities that original data collection endeavours entail, ensuring data are accurate, credible, legitimate and consistent across countries and issue areas over time. Despite the positive reception of our work and the success we have enjoyed for producing reliable and valid data, we are increasingly wondering whether we are on track to achieve the impact we aim for.

Our logic for producing actionable governance data has traditionally revolved around a simple set of causal assumptions that were more often articulated internally than publicly: the data generated about governance and corruption issues would be taken up by domestic change agents to design evidence-informed governance reforms, yielding—over time—improvements in transparency and a reduction in public corruption.

However, recent research about the uptake and impact of governance data has cast doubts about the accuracy of this logic (Custer et al. 2016; for a different view, see Kelly and Simmons 2015). Key questions remain unanswered about when and how governance data are useful to domestic actors to indeed spur reform. How can we (and other external actors) make sure data are

meaningful and useful to the domestic actors actually responsible for governance reform? How can we foster and support the uptake of the data? What concepts and proxies should be measured, given limited resources and imperfect methodologies? What principles and benchmarks should be used to evaluate the impact of our data? How do we evaluate their effectiveness?

While the debate traditionally centres on methodological questions and particular approaches to enhancing the reliability and validity of the data, we think these questions are primarily of ontological nature (Hudson 2014). They were thus a key driver for Global Integrity revamping its organisational strategy in the first half of 2015 (Global Integrity 2015b). Our new strategy is built around an explicit theory of change (see Valters 2015), and we lay out the assumptions and mechanisms we believe our programmes should address in order to achieve impact. Being open and precise about what we know and what we do not know, and inquiring not just into methodological questions but also into the logic of governance reform writ large, will enable us to iteratively sharpen our impact and effectiveness in order to make a meaningful difference. We remain committed to supporting progress towards open governance by collecting high-quality reliable and robust data, but we have started to put a stronger emphasis on questions about data use and usability to better understand how data can support domestic stakeholders in driving governance reform in their decision-making ecosystems (Global Integrity 2016a).

Inquiring into these questions is as much about the work and impact of other organisations in the field as it is about Global Integrity's work. Attempting to find satisfactory answers is no easy feat and we encourage others to join us on this journey, sharing their insights and strengthening our collective thinking. We hope our account will be of interest to readers, will stimulate debate and will provide at least partial insights in response to some of the challenges and questions others have raised in the first part of this volume.

A Brief History of the Global Integrity Report

In the early 2000s, questions started to emerge about the validity and usefulness of the existing measures of corruption. The Worldwide Governance Indicators and Transparency International's Corruption Perceptions Index, for example, were acknowledged to contribute to raising awareness, communicating broad challenges and providing (imperfect) comparable benchmarks. At the same time, practitioners agreed that there was a need for improved

diagnostics to support policymakers, civil society advocates and other stakeholders in using the information and responding in targeted and effective ways to the problems of governance and corruption.

A common criticism of popular external measures of corruption at the time was that they were "extremely broad", and not actionable (Global Integrity & United Nations Development Programme 2008). The findings of these multi-country large-sample composite indices allowed comparisons of scores across countries and over time but did not provide insights into specific areas of governance reform needs and challenges existing within established anti-corruption frameworks. Perception indices were largely considered to fail to help governments and advocates identify laws they could prioritise to make improvements in a timely way.

Another key challenge was the limited focus of existing research on legal provisions rather than inquiring into whether they were successfully implemented on a day-by-day basis, bringing about the intended results. Reform-minded practitioners rightly pointed out that effectiveness of implementation was the key to judging the effectiveness of a set of laws.

Last but not least, we identified yet another unresolved challenge. In order for citizens to participate and advocate for better governance, they needed access to the transparency and accountability measures and mechanisms described in the law. Could citizens effectively access such information and use the mechanisms at hand? Responding to these challenges and identifying an entry point to enhance the data landscape, Global Integrity set out to generate actionable and action-worthy qualitative data, drawing upon around 325 indicators to assess the anti-corruption safeguards of countries.

Our then implicit theory of change for the Global Integrity Report—our model of why and how change happens and our role in contributing to the potentially desired outcome—can be summarised as follows:

1. In order to meaningfully enhance a country's anti-corruption framework in pursuit of more integrity, decision-makers have to make better evidence-based policy decisions.
2. Global Integrity supported (and possibly catalysed) these efforts by producing actionable and action-worthy evidence, and collecting original data that were locally sourced, factual and comparable across countries.
3. This information—combined with a vivid narrative prepared by journalists in the form of accompanying Reporter's Notebooks—was made available cost-free to inform and empower citizens, activists, donors, businesses and governments to take action in their respective countries.

The logic underlying this model can be explained by understanding Global Integrity's genesis. Conceived in 1999 as a pilot project by the Center for Public Integrity (CPI), an investigative reporting non-profit organisation in Washington, D.C., Global Integrity set out to test new ways of investigating and assessing corruption and to find out how governments in particular were addressing it. Success of its initial pilots soon led to Global Integrity spinning off from the CPI as a separate organisation focused on tracking governance and corruption trends worldwide. The legacy was a firm rooting of our methodological approach in a blend of social science and journalism, and trusting the power of reporting to affect change in the world.

Measuring the Opposite of Corruption

Our indicators are based on a simple yet powerful concept. Rather than trying to measure actual corruption, which experts consider virtually impossible (Arndt and Oman 2006), Global Integrity qualitatively assesses the *opposite* of corruption, that is, integrity provisions that allow for citizens and businesses to have access to a country's government and define their ability to monitor its behaviour, to seek redress and advocate for improved governance (Camerer 2006). Its indicators break down this "access" into a number of categories and questions, ranging from inquiries into electoral practices and media freedom to budget transparency and conflict of interest regulations. We unpack these concepts by not only looking at the laws or institutions that are "on the books" but also by assessing their implementation and enforcement through indicators assessing staffing, budget support, political independence and citizen access to the most important anti-corruption mechanisms.

Perhaps the most critical element introduced by Global Integrity is the way the indicators break down the components of a country's anti-corruption system, looking into checks and balances within the executive, legislature and judiciary, as well as other relevant areas such as independent audits and civil service whistle-blower protection and so on. This allowed the indicators to clearly pinpoint and identify the most critical vulnerabilities—or weakest links—in the system, with the report serving as an actionable roadmap for decision-makers. Relatedly, and since the indicators sought to be based not on perceptions but on relevant current references, they could be used as an evidence-based tool to aid decision-making. The premise was that a policymaker would have access to more than a general and unhelpful score about the extent of the corruption problem in her country but instead could: (a) dive deep into specific weak links in the existing framework, identifying

failing mechanisms or provisions not deterring corruption, and (b) find evidence-backed data points to support informed policy design.

Another innovation of the Global Integrity Report was its introduction of the "implementation gap". The measure was originally introduced by Global Integrity co-founder and then Executive Director Nathaniel Heller, primarily as a means to generate increased media attention around the annual Global Integrity Report. Over time, it became clear that the implementation gap was actually a highly useful practical insight for policymakers and advocates alike. The gap is calculated at the country score level by deducting the "de facto" or implementation score from the "de jure" score. The resulting score provides an important reference about the extent to which legal provisions and mechanisms are or are not actually being implemented (Nadgrodkiewicz et al. 2012).

A New Methodology Is Born

Our methodological approach was built around the idea of integrating social science and journalism by relying on in-country teams of independent researchers and journalists to report on the *de jure* and the *de facto* reality of corruption and governance and then constructing our indicators based on best-practice benchmarks with a strong emphasis on concept validity, quality control and a double-blind peer review process. The Global Integrity Report follows an expert-based assessment methodology, aggregating 325 indicators to generate an "Integrity Scorecard".

The indicators are designed to provide a qualitative snapshot of the anti-corruption safeguards in a particular country and to "score" the institutional framework that exists at the national level with a view to promoting public integrity and accountability, preventing abuses of power, and implementing and enforcing regulations where needed. Scoring the indicators (and for a long period of time ranking countries on the basis of their overall scores) provided a helpful shorthand way of structuring and presenting the qualitative data obtained through our research as a snapshot of a particular situation. It was deemed an important logical step in our endeavour to drill deeper into particular mechanisms, all the while staying away from writing and publishing non-comparable case studies.

The original indicator selection was based on a detailed and comprehensive literature review together with other sources of input such as interviews with experts and state-of-the-art thinking in the field. The indicators examine three concepts:

1. The *existence* of public integrity mechanisms: the laws, regulations and agencies/entities or equivalent mechanisms in place in a particular country.
2. The *effectiveness* of these mechanisms: aspects of public integrity such as protection from political interference, appointments that support the independence of an agency, professional full-time staff and funding, independently initiated investigations and the imposition of penalties.
3. Citizen *access* to these mechanisms: the availability of public reports and information to citizens within a reasonable time and at a reasonable cost.

Until 2013, when the most recent iteration of the Global Integrity Report was published, the indicators were organised in six main categories:[1] (a) Non-governmental Organizations, Public Information and Media; (b) Elections; (c) Government Conflicts of Interest Safeguards and Checks and Balances; (d) Public Administration and Professionalism; (e) Government Oversight and Controls; and (f) the Anti-Corruption Legal Framework, Legal Impartiality and Law Enforcement Professionalism (Global Integrity 2011).

Within each category, there are a number of subcategories and eventually specific indicators—measurable and conceptually valid proxies—assessing the state of the subcategories by looking at all three concepts for each subcategory where feasible. Each indicator is scored by an in-country researcher and substantiated as far as possible with relevant accurate updated references and additional comments which come from interviews with key authoritative sources, such as experts and stakeholders on the ground, document reviews, news articles, studies and so on. At first, it was optional for researchers to provide additional commentary and clarifications to support their score choices. However, when they were provided the results turned out to be particularly useful in capturing specific nuances of situations and country contexts.[2] In later iterations, the option of providing comments became a hard and fast requirement to allow for 100% transparency of the reasons why a researcher chose a specific score in response to specific scoring conditions.

All the indicators are scored on a scale of 0 (worst) to 100 (best) and there are two types of indicators: "in law" and "in practice". "In law" indicators provide an objective assessment of whether certain legal codes, fundamental rights, government institutions and regulations exist. These *de jure* indicators are scored with a simple "Yes" or "No", with "Yes" receiving a score of 100 and "No" receiving a zero. The "in-practice" indicators address *de facto* issues such as implementation, effectiveness enforcement and citizen access. As they usually require a more nuanced assessment, they are scored along an ordinal scale from 0 to 100 with possible scores at 0, 25, 50, 75 and 100.

Global Integrity reviews the scores for reliability and overall quality. In addition, a set of in-country peer reviewers check the scores and related references to identify possible bias, inaccuracies or omissions. The researchers and peer reviewers do not know each others' identities. The double-blind nature of the peer review process guarantees feedback free of considerations associated with personal bias regarding data collection and the scoring of the indicators to avoid the risk of a peer-influenced consensus. To further minimise the effect of subjective perceptions and to maximise inter-coder reliability, Global Integrity provides researchers and peer reviewers with clearly defined scoring criteria for every single indicator. These anchor each indicator to a predefined set of criteria.

To produce a country's aggregate scorecard, a simple aggregation method is used. Each indicator score is averaged within its parent subcategory, which produces a subcategory score. The subcategory score is in turn averaged with the other subcategory scores into a parent category score. The category scores are averaged to produce an overall country score.[3] The Global Integrity Report clusters countries into five performance "tiers" according to their overall aggregate scores.[4]

Global Integrity's approach is thus characterised by an effort to be fully transparent about the methodology employed, the scoring choices made, the scoring rationales and the sources consulted to enhance the robustness and credibility of our findings (Global Integrity 2011).

Evolution of the Expert Assessment Methodology

Important methodological developments have taken place since the first iteration of the report in terms of the research process and its underlying logic, the focus and emphasis of the data collected, quality control measures and not least the publication strategies for the GIR and other reports.

The first big change was simple and involved country coverage. The first few iterations of the GIR covered nearly 70 countries annually. Starting in 2010, the report covered about half of that number 1 year and the other half the following year. Given that score changes for countries were usually found to be minimal from 1 year to the next, we decided that resources could be better used to launch new indicator-based projects.

In 2011, another change was introduced with the GIR no longer including the Global Integrity Index, which had previously ranked countries by their overall scores. This was due to the decreased coverage of the Report, which reduced the utility of the Index, and also due to a belief that indices can be

effective as dissemination tools but not necessarily as policy-making tools. "Country rankings are too blunt and generalised to be 'actionable' and inform real debate and policy choices. Sure, they can put an issue on the table, but that's about it", as Global Integrity explained at the time (Heller 2011a).

Another important change affected Global Integrity's take on its expert assessment methodology in 2012. A significant review and reform of its methodology started, leading to major changes in both the design of the indicators and the processes utilised to gather data in the field, including revising and strengthening the quality control processes conducted by Global Integrity.

The Global Integrity Report indicators did not undergo significant changes beyond category re-organisation and minor language changes until 2013, when the Report was put on pause awaiting a revamp in the near future. However, we further expanded into new projects such as the Africa Integrity Indicators (Global Integrity 2016a), the Money, Politics and Transparency (MPT) project (Global Integrity 2015a) and the State Integrity Investigation (SII) (Global Integrity 2015a). We have continued to enhance our expert assessment methodology, improving its processes and mechanisms regarding indicator design, the data collection process and the data quality control process.

Indicator Design

The tension between prescriptive and "fuzzy" approaches is an important factor when defining an indicator and the benchmarks to be utilised to operationalise it (the conditions that must be found in the field to select a given score or answer). Global Integrity has traditionally avoided rigid prescriptive approaches, understanding the disadvantages of "one size fits all" measurements. On the other hand, indicators that fail to explicitly unpack and define certain concepts become too fuzzy for consistent coding—and therefore diminish the value of a final product that proposes to offer users practical value. In response, the wording of the indicators has become much more specific and the resulting data more consistent and actionable (see World Bank 2016b).

Evidently, a diversity of points of view exists about any of the issues assessed by the indicators, even more so when it comes to defining concrete benchmarks for the scoring criteria. In construing the indicators, Global Integrity makes an effort to work with commonly accepted views, best practice and international benchmarks, but understands that in many cases there will still be alternative views. For this reason, our indicators, including the questions,

benchmarks and scoring conditions, together with selected scores, narratives and the references consulted are made publicly available so that users can ultimately determine for themselves whether the data prove useful for their measurement objectives.

This leads to another key issue: the evidence-based nature of Global Integrity's contribution to the field. The effort to define questions and scoring conditions in a more concrete and explicit way comes attached to an obligation to also provide more concrete and explicit research and facts to back up scoring choices. If it were not for the narratives provided by experts, there would be a visible mismatch between the scoring conditions for granular questions and generic or otherwise vague responses, which could generate the impression that a score is not based on detailed research and therefore result in a credibility risk. Indeed, this means taking the traditionally conceived expert assessment methodology one step further towards what we call a fact-based expert analysis.

Furthermore, the practical realities of data collection also impact the definition and construction of indicators. In some cases, this means that certain indicators may not be the right tools to assess certain issues, not because of a flawed concept but because of the inability of researchers to meaningfully document their findings given the nature of the subject matter and/or time and resource constraints. For example, after several rounds of research it became evident that the information necessary to score one indicator in the Global Integrity Report was not publicly available in most countries. In addition, the perceptions of the experts often consulted about the issue were not based on first-hand information and for that reason tended to be vague. In such cases—and even though these indicators may still try to answer legitimate questions—Global Integrity adopted a policy of avoiding their inclusion so as to not compromise the rigour and reliability of the research.

Another change we experimented with during the production of the MPT data set was to include semi-structured open questions in the survey to allow for yet more opportunities to report context-specific information. Obviously, we exempted the indicator question from influencing the numerical score. As a result, and despite this seemingly minor change to the methodology, researchers were able to provide case study-like material in the scorecard to illuminate the context of particular indicator categories.

The Data-Gathering Process

Under the initial methodological approach as pursued during the first years of the GIR, researchers were allowed to justify their score choices simply by

providing a list of references which included all the sources consulted. Commentary with an explicit rationale to justify the score choice was optional and the researchers did not explain many of their scores.

Coming to understand the value of fact-based narratives explaining why a particular score was chosen, we updated the approach to requiring researchers to always provide an explicit rationale explaining each and every score choice. In addition, we formalised the requirements regarding the sources to be consulted. As a minimum, each researcher would have to consult and list at least three relevant accurate current sources of information. These changes had two main inspirations:

1. Explicit reasoning behind the score choices easily allowed identification of instances where the researcher's interpretation of the indicator was not consistent with the scoring criteria and/or showed bias or inaccuracies.
2. Users of the data had repeatedly expressed that, in addition to the scores, they found great value in indicator-specific details and explanations provided by the researchers.

Another modification relates to the use of expert opinions. The original methodology allowed an expert's opinion on an issue to be cited as the source to back a score choice. However, given the usually small number of experts consulted for any given indicator and the fact that experts sometimes disagree, researchers were subsequently asked to go beyond opinions (as far as possible), by verifying the statements of experts against documented facts (whenever feasible), by pressing experts for more detailed reasoning for their statements, or by obtaining their reactions to contrasting opinions. Additionally, researchers were required to provide a variety of sources and base their assessments on a comprehensive set of sources, including, for example, interviews, desk research into studies and news coverage and as much field observation as possible.

The scoring scale was also slightly modified in order to better capture the realities observed in the field. In the Global Integrity Report all "in law" indicators were scored with either "Yes (100)" or "No (0)". If a law existed in a country but only partially met the conditions of the scoring criteria, it was scored a No (0) as if it did not exist. Our new projects include a "partial" as part of the benchmark and allow for an intermediate score (50) in order to better reflect situations where a law exists.

In the same way that challenges in the data collection process can impact the final quality of the indicators, regardless of how careful the conceptualisation and the methodological design is, the quality control and project

management processes can also heavily impact data quality. For this reason, these two processes have also undergone significant improvements, including among other things individual training for the peer reviewers, new manuals to ensure standardised quality control by the Global Integrity team, more frequent communication with in-country researchers as they conduct the fieldwork and discussions among the team to identify possible consistency issues across specific indicators.

To further enhance the quality control process and to experiment with enhanced outreach for increased uptake and use of the data, in 2016 for the first time, we started provisionally publishing the Africa Integrity Indicators data. Following the provisional publication, stakeholders, including governments and civil society alike, have 2 months to review the information and request score changes if they can provide factual documented information justifying the need for them.

New Data Sets in an Expanding Field

Many data users including governments and civil society organisations continue to rely on our data either directly or indirectly through its incorporation in the Worldwide Governance Indicators. Some of the best-known users include the Millennium Challenge Corporation (MCC), the World Bank and the Australian Department of Foreign Affairs and Trade (DFAT, then AusAid). Private sector users include, for example, Moody's Investors Service and TRACE International. Despite acknowledgement and public support from important users praising our data, few funders were willing or able to consider financing a globe-spanning continuation of the Global Integrity Report. Putting the GIR on hold in 2011 was therefore primarily a reaction to a fundraising challenge Global Integrity was facing.

At the same time, there seemed to be an increased interest in funding regional and domain-focused work and this aligned with the organisation's interests. Global Integrity had launched a number of smaller projects, such as sector-specific assessments (e.g. an assessment of transparency in the justice system in Guatemala or access to information laws in the largest Mexican cities) and subnational assessments (assessing anti-corruption frameworks and their application at the subnational level in a small number of countries in Latin America, Africa and Asia) starting in 2008 (Global Integrity 2015c).

In this context, other partners emerged with an interest in taking the data collection forward into new areas. In 2012, Global integrity and the Mo Ibrahim Foundation went into partnership to launch a new piece of research,

the Africa Integrity Indicators (AII) (Global Integrity 2016b), with indicators focusing partly on integrity issues and partly on human development matters covering all 54 African countries for 5 years. The partnership was mutually beneficial. It allowed Global Integrity to continue collecting original data on governance questions through in-country researchers throughout Africa, and it allowed the Foundation to draw on specific in-practice human development indicators to fill gaps in their Ibrahim Index of African Governance (IIAG). As of mid-2016, Global Integrity is gathering data for the fifth round of research. The partnership is envisioned to continue for another 5 years.

The AII project has resulted in a number of exchanges with government representatives. These usually engage with Global Integrity in one of the following three ways: (a) out of their own initiative—if they have questions or suggestions about the data; (b) by our invitation, this mode has increased sharply as we have started to provide access to the provisional data, which is published by the end of March each year; and/or (c) by the encouragement of the Millennium Challenge Corporation or the World Bank (or other entities) when they use our data to inform their decision-making in relation to their country programmes and subsequently advise governments to consult with us directly to understand where the data come from, how the research was conducted and what our findings are at the indicator level.

In 2012, we started collecting data for the first round of the State Integrity Investigation, which assesses the corruption risk in 15 areas of government operations across all 50 US states. The project was a collaboration between the Center for Public Integrity, International Public Radio and Global Integrity. In 2015, we published a second round of research, this time in sole partnership with the Center for Public Integrity. In contrast to the last few iterations of the GIR and the AII project, the SII results were published as a ranking (Global Integrity & Center for Public Integrity 2015a, b). The data were accompanied by an in-depth news report for each state. During the second iteration we counted more than 2,000 media mentions with at least 12 high-level representatives of state governments publicly referencing the results and arguing for strengthened ethical controls in their respective states.

We have also launched another data-gathering project in collaboration with the Electoral Integrity Project (EIP) and the Sunlight Foundation to assess campaign finance rules and the extent to which they were implemented in practice in 54 countries during 2014 and 2015 (Global Integrity 2015a). The Money, Politics and Transparency (MPT) project was conceived to link indicator-based data collection with academic research, advocacy and norm-setting practice. It resulted in a number of government organisations enquiring about best practices or discussing reforms they had in mind.

From Best Practice to Best Fit

Despite this exciting anecdotal evidence of excellent uptake of our work by practitioners from within government and civil society, we know it is neither the provision of data nor the data themselves that inspire change or lead to impact. Data on their own constitute just one factor among many others that enable uptake and contribute to supporting reform. Evidence on the dynamics and the precise combination of factors enabling and driving change is still difficult to find (see Kleinfeld 2015) and the anecdotal evidence we have collected about where and when uptake of our data happens and how it translates into change is (at best) patchy.

While we are delighted with all the governments that are making the effort to reach out and engage, and elated with every official referencing of our data to argue for stronger reforms, we are keenly aware that the vast majority of governments and officials are neither aware of evidence-based governance reform and nor do they make it a priority.

Achieving greater clarity about the use and usefulness of governance data under varying conditions and bolstering our collective understanding of how data producers can contribute to making governance assessments more impactful has thus become a core priority for Global Integrity.[5] We are pursuing this priority by experimenting with and exploring our governance assessments on the basis of a revised strategy with adaptive learning at its core and through our participation in the Governance Data Alliance (GDA).[6]

In 2015, we set out to rethink our then implicit theory of change with a view to crafting an organisational strategy that would guide our work thenceforth.[7] The aim was to make our assumptions about how and why change happens explicit, and to commit ourselves to periodically revisiting these assumptions and our programme logic to ensure our efforts are on track and responsive to what we are learning. Our strategy centres on the concept of adaptive learning—a structured approach to learning by doing—as the key mechanism in our theory of change, aimed at supporting progress towards more open and accountable governance in countries and communities in pursuit of better development outcomes around the world (Global Integrity 2016a). [8]

Our strategy is based on three key insights about how countries move towards governance reform. First, progress towards more open governance is inherently political. It follows that our approach must be about understanding and engaging with the politics of governance reform; it cannot be purely technical and removed from the political realities which domestic actors face and experience. Second, within a given governance context the primary role

in driving governance reform is played by domestic actors. Therefore, the task of defining problems and strategising about solutions resides primarily with domestic champions; external actors play a supporting role. Third, there is no universal blueprint for governance reform. This means reformers must engage in a process of trying, learning and adapting their approach to governance reforms that work in their contexts. It also means that we must support the process in contextually appropriate ways.

Flowing from these insights about how reform happens and the role external actors play throughout the process, our theory of change holds that people and organisations can increase their impact and effectiveness by putting adaptive learning into practice, implementing a structured process of learning by doing. In particular, this means:

> taking action in pursuit of an objective; monitoring to collect data on the effectiveness of that action; using that data to reflect on the approach taken; and adapting the approach for future iterations. More concisely, adaptive learning is a continuous process of theoretically informed data-driven reflection and adaptation aimed at improving impact and effectiveness. Adaptive learning—when it engages the political dynamics of development, strengthens reform coalitions and informs political action—provides a means of navigating complexity, closing the gap between policy commitments and implementation, addressing delivery challenges and delivering results. Adaptive Learning starts with the definition of specific problems in particular contexts and supports the search for locally appropriate solutions. (Global Integrity 2016a)

The approach has wide-ranging implications for our programmes, including our work on governance measures. The core hypothesis is that the provision of relevant and actionable data can facilitate politically engaged adaptive learning and data-driven reflection on how policies are playing out in practice and on how different actors might work together, navigating and shaping the political landscape to drive progress towards more open and effective governance. In other words, engaging in data collection projects is as much about supporting reformers adaptively learning to engage with each other around data in politically informed ways to reach consensus, create coalitions for reform and try out context-specific solutions as it is about describing and measuring governance phenomena.

Spelling out clearer assumptions about how governance assessments can contribute to more open and effective governance has practical implications along a number of dimensions. Exploring these dimensions, trying them out in practice, learning about what is working and for what reasons and reflecting and adapting our approach is a core objective that we pursue under the

heading of "Governance Assessments 2.0" (Tonn 2016). This effort combines our thinking about the importance of adaptive learning with our strength and expertise in producing reliable and robust governance and anti-corruption measures.

One dimension that we are rethinking involves the role that external "best-practice" benchmarks should play vis-a-vis locally defined "best-fit" benchmarks. Traditionally, and in line with the best-practice approach, Global Integrity has strongly relied on political science concepts to guide our thinking around best-practice benchmarks and their applicability to our measurements. Benchmarks are usually defined by academia or by standard-setting organisations such as the Organisation for Economic Co-operation and Development (OECD) or the World Bank, or in other cases by global non-governmental organisations such as the International Institute for Democratic and Electoral Assistance (International IDEA) or the International Budget Partnership (IBP). While we have obviously taken steps to not blindly rely on such benchmarks—by double-checking their relevance and applicability in consultations with stakeholders on the ground—our reliance on best-practice benchmarks has nonetheless meant that we measure reality against idealised concepts.

Moving from best practice to best fit will require a more nuanced understanding of the pathways governance reform can take in particular situations, and greater insights into which benchmarks are useful to the actors driving reform. Furthermore, it may be necessary for the field to collectively accept the value of best-fit solutions (and institutions) while acknowledging that we might have to forego some cross-country comparability in order to generate reliable context-specific data.

A corollary is the absolute need to coordinate much more closely with domestic stakeholders. Aligning assessments with local priorities and making sure that we measure what counts (for them) is a task that can only be completed if and when our assessments build on buy-in by these actors and draw on the deep insights and understanding they bring to the table. This logic arises as a direct consequence flowing from our strategy.

To be clear, we are by no means the first organisation to think about ways to assess locally defined priorities. Examples include the work done by MacGinty and Firchow (n.d.) on bottom-up peace indicators, the participatory tracking project by the Social Observatory (2013), US Agency of International Department's Good Governance Barometer (USAID and Family Health International 360 2015) and DataShift's work on citizen-generated data (Gray et al. 2016), among many others. We are keen to build on this work and learn from existing projects and tried strategies as we attempt

to integrate these concepts in a way that will allow for both meaningful insights and reliable robust data.

Another dimension we have grappled with is whether and how to measure function rather than form (Hudson 2014). At Global Integrity, we have sought to assess intermediate outcomes in addition to form, focusing on the implementation effectiveness of particular institutions. However, we are aware that we are far from this goal and that there is an attribution gap we cannot bridge with our indicators (Heller 2011b). There is no question: it is methodologically challenging to zero in on function and relinquish the more easily quantifiable scoring conditions utilised to assess form. It is encouraging, however, that various ideas are emerging to hone in on measuring outcomes in a way that is meaningful to governments and reformers alike (see, e.g. Andrews et al. 2010; World Bank 2016a). Our approach to further exploring this question will consist of utilising the principles of adaptive learning to engage with stakeholders to understand which functions they deem important. Working backwards, we will support them to evaluate their ecosystem of data and data usage to temporarily measure the forms they believe can help them achieve their aims.

Conclusion

In this chapter, we have laid out how and why asking tough questions about the usefulness of data, methodological rigour, reliability and validity has resulted in important improvements to our methodological approach and how this process has driven the evolution of governance indicators at Global Integrity. We have described the beginnings and the underlying rationale of our work, how we have reacted to changing priorities and how we have improved upon our (at first) implicit theory of change.

We believe there is a need to collectively learn, across the field, about how we can strengthen the mechanisms leading to uptake and reform so that governance assessments can live up to their promise and indeed deliver impact. The shift towards increased usefulness at the country level will bring new challenges and methodological trade-offs to the fore. We are certain, however, that these challenges can be resolved.

Reliability, validity and analytical rigour will continue to rank high on our list of methodological guidelines, but it is time to focus on locally sourced and domestically legitimate benchmarks, prioritising function over form and emphasising the utility of context-specific data to support reform champions as they experiment, learn and adapt their way towards reforms that work in their contexts.

Making our thinking explicit—by laying out the value of adaptive learning and shifting the emphasis towards exploring and better understanding the use and usefulness of data before engaging in data collection projects—will help us to meaningfully address and grapple with the challenge of how we can best support domestic actors in driving governance reform, based on local priorities, at home.

Notes

1. The full list of indicators is omitted given the large number. However, they are available at https://www.globalintegrity.org/downloads/. Small modifications were made to both the categorisation and wording of the indicators during the various iterations of the Global Integrity Report.
2. The researchers are highly qualified and experienced individuals in each country, usually from the fields of political science, law, journalism or academia. They are required to have significant experience in anti-corruption issues and cannot have had recent contractual engagements with the government.
3. The categories are equally valued, even though some categories are derived from a lengthier series of sub-indicators/questions than others. Similarly, the subcategories are equally valued within their parent category.
4. The scores are: Very Strong (90+), Strong (80+), Moderate (70+), Weak (60+), Very Weak (<60).
5. While there are a number of organisations thinking along the same lines (see, e.g. Reboot 2015; Custer et al. 2016), the vast majority of data-producing organisations still operate, at least publicly, on the assumption that the mere creation of information coupled with basic dissemination models will lead to change.
6. The alliance is a consortium of governance data producers, users and funders working together to strengthen the production, use and impact of governance data. It is currently hosted by the Results for Development Institute and was initially convened by Global Integrity co-founder and then Executive Director Nathaniel Heller in 2014. A major contribution that the Alliance has made is commissioning a first-of-its-kind user survey to explore the use of governance data, conducted by AidData. Among other findings, the report showed data are most useful if and when they are aligned with country priorities and that the political context plays a role in uptake and usefulness. The results—as noted throughout this chapter—accentuate and corroborate our longstanding unease about the real impact of our (and other organisations') data.
7. Executive director and co-founder Nathaniel Heller left the organisation at the end of 2014 and Alan Hudson took the helm of the organisation in early 2015. Hazel Feigenblatt left in mid-2015.

8. Open and accountable governance has always been at the heart of what Global Integrity does. As set out earlier, our initial hypothesis—or theory of change—revolved around providing data and information for change agents to catalyse reform. Moreover, while we revised various aspects of that logic over time, it was an internal and rather unsystematic undertaking. Starting in 2015, we have pivoted to hypothesising about the value of adaptive learning as the key mechanism in our theory of change and we have established mechanisms to transparently explore this theory and reflect on our actions and learning, committing to showcase both our progress and our failures and to inviting feedback and debate from outside the organisation.

References

Andrews, M., Hay, R., & Myers, J. (2010). *Governance Indicators Can Make Sense: Under-Five Mortality Rates are an Example* (Harvard Kennedy School. Faculty Research Working Paper Series, RWP 10-015). Cambridge, MA: Harvard Kennedy School. Retrieved from https://research.hks.harvard.edu/publications/getFile.aspx?Id=541

Arndt, C., & Oman, C. (2006). *Uses and Abuses of Governance Indicators*. Paris: Organisation for Economic Co-operation and Development (OECD). Retrieved from http://www.la.utexas.edu/users/chenry/polec/2006/oecd/AE795835C8392A8111572211048C64BBAF3DA2573E.pdf

Camerer, M. I. (2006). Measuring Public Integrity. *Journal of Democracy, 17*(1), 152–165.

Cooley, A. (2015). The Emerging Politics of International Rankings and Ratings: A Framework for Analysis. In A. Cooley & J. Snyder (Eds.), *Ranking the World: Grading States as a Tool of Global Governance* (pp. 1–38). Cambridge: Cambridge University Press.

Custer, S., Takaaki, M., Sethi, T., Latourell, R., Rice, Z., & Parks, B. (2016). *Governance Data: Who Uses It and Why? Governance Data Alliance*. Retrieved from http://www.r4d.org/govdatareport

Global Integrity. (2011). *The Global Integrity Report: 2011 Methodology White Paper*. Washington, DC: Global Integrity. Retrieved from https://www.globalintegrity.org/wp-content/uploads/2013/08/2011_GIR_Meth_Whitepaper.pdf

Global Integrity. (2006–2013). *Global Integrity Report*. Washington, DC: Global Integrity. Retrieved from https://www.globalintegrity.org/research/reports/global-integrity-report/

Global Integrity. (2015a). *Money, Politics, and Transparency*. Washington, DC: Global Integrity. Retrieved from https://data.moneypoliticstransparency.org/

Global Integrity. (2015b). *Learning to Open Governance. Strategy Document*. Washington, DC: Global Integrity. Retrieved from https://www.globalintegrity.org/wp-content/uploads/2015/07/GlobalIntegrityLearningtoopengovernanceStrategy.pdf

Global Integrity. (2015c). *Sub-national and Sector Datasets*. Washington, DC: Global Integrity. Retrieved from https://www.globalintegrity.org/downloads/

Global Integrity. (2016a). *Learning by Doing: Our Action Plan for Open Governance.* Washington, DC: Global Integrity. Retrieved from http://www.globalintegrity.org/wp-content/uploads/2016/09/Global-Integrity-Learning-Plan-6th-September-2016.pdf

Global Integrity. (2016b). *Africa Integrity Indicators*. Washington, DC: Global Integrity. Retrieved from http://aii.globalintegrity.org/

Global Integrity & Center for Public Integrity. (2015a). *How Does Your State Rank for Integrity?* Washington, DC: Global Integrity, Center for Public Integrity. Retrieved from https://www.publicintegrity.org/2015/11/09/18822/how-does-your-state-rank-integrity

Global Integrity & Center for Public Integrity. (2015b). *State Integrity Investigation*. Washington, DC: Global Integrity, Center for Public Integrity. Retrieved from https://www.publicintegrity.org/accountability/state-integrity-investigation/state-integrity-2015

Global Integrity & United Nations Development Programme (UNDP). (2008). *A User's Guide to Measuring Corruption*. Oslo: UNDP Oslo Governance Centre. Retrieved from http://www.undp.org/content/undp/en/home/librarypage/democratic-governance/anti-corruption/a-users-guide-to-measuring-corruption.html

Gray, J., Lämmerhirt, D., & Bounegru, L. (2016). *Changing What Counts. How Can Citizen-Generated Data and Civil Society Data be Used as an Advocacy Tool to Change Official Data Collection?* DataShift and Open Knowledge. Retrieved from http://civicus.org/thedatashift/wp-content/uploads/2016/03/changing-what-counts-2.pdf

Heller, N. (2011a). *Why We Killed the Global Integrity Index*. Washington, DC: Global Integrity. Retrieved from http://www.globalintegrity.org/2011/05/post-792/

Heller, N. (2011b). *Outputs Versus Outcomes in Open Government*. Washington, DC: Global Integrity. Retrieved from http://www.globalintegrity.org/2011/09/outputs-and-outcomes-opengov/

Hudson, A. (2014). *Measuring Governance: What's the Point?* Washington, DC: Global Integrity. Retrieved from http://www.globalintegrity.org/2014/06/measuring-governance-whats-the-point/

Hudson, A. (2015). *The Marketplace of Ideas: From "External Assessments" to Country-Level Learning*. Washington, DC: Global Integrity. Retrieved from http://www.globalintegrity.org/2015/05/the-marketplace-of-ideas-from-external-assessments-to-country-level-learning/

Kelly, J., & Simmons, B. (2015). Politics by Numbers: Indicators as Social Pressure in International Relations. *American Journal of Political Science, 59*(1), 55–70.

Kleinfeld, R. (2015). *Improving Development Aid Design and Evaluation. Plan for Sailboats Not Trains*. Washington, DC: Carnegie Endowment for International Peace. Retrieved from http://carnegieendowment.org/2015/03/02/improving-development-aid-design-and-evaluation-plan-for-sailboats-not-trains-pub-59159

MacGinty, R., & Firchow, P. (n.d.). Everyday Peace Indicators: Capturing Local Voices Through Surveys. *Shared Space: A Research Journal on Peace, Conflict and Community Relations in Northern Ireland, 18*, 33–39. Retrieved from http://www.community-relations.org.uk/wp-content/uploads/2014/12/Everyday-Peace-Indicators-Capturing-local-voices-through-surveys.pdf

Malito, D. V. (2014). *Measuring Corruption Indicators and Indices* (Robert Schuman Centre for Advanced Studies Research Paper 2014/13). Florence: European University Institute. Retrieved from http://dx.doi.org/10.2139/ssrn.2393335

Nadgrodkiewicz, A., Nakagaki, M., & Tomicic, M. (2012). *Improving Public Governance: Closing the Implementation Gap Between Law and Practice*. Washington, DC: Center for International Private Enterprise and Global Integrity. Retrieved from http://www.cipe.org/publications/detail/improving-public-governance-closing-implementation-gap-between-law-and-practice

Reboot. (2015). *Using Data to Influence Government Decisions. Opportunities and Challenges for User-Centered Design to Improve Governance Data Impact*. New York: Reboot. Retrieved from http://reboot.org/wordpress/wp-content/uploads/2015/07/Reboot_Using-Data-to-Influence-Government-Decisions_2015.pdf

Social Observatory. (2013). *Participatory Tracking Data Collection*. Washington, DC: World Bank. Retrieved from http://www.worldbank.org/en/news/video/2013/12/29/the-social-observatory-participatory-tracking

Tonn, J. (2016). *Toward Governance Assessments 2.0*. Washington, DC: Global Integrity. Retrieved from http://www.globalintegrity.org/2016/03/toward-governance-assessments-2-0/

US Agency of International Development (USAID) & Family Health International 360 (FHI 360). (2015). *Guide to the Good Governance Barometer*. Washington, DC: USAID & FHI 360. Retrieved from http://pdf.usaid.gov/pdf_docs/PA00KJMP.pdf

Valters, C. (2015). *Theories of Change: Time for a Radical Approach for Learning in Development*. London: Overseas Development Institute. Retrieved from https://www.odi.org/sites/odi.org.uk/files/odi-assets/publications-opinion-files/9835.pdf

World Bank. (2016a). *Draft World Development Report 2017*. Washington, DC: World Bank. Retrieved from http://www.worldbank.org/en/publication/wdr2017

World Bank. (2016b). *Actionable Governance Indicators and AGI Data Portal*. Washington, DC: World Bank. Retrieved from https://www.agidata.org/site/ and http://data.worldbank.org/data-catalog/actionable-governance-indicators

13

Developing an Indicator of Fiscal Sustainability for Africa

Moses Obinyeluaku

Introduction

The special nature of oil revenue complicates evaluation of the macro-fiscal stance of oil-producing countries. An accurate assessment of this issue can be obscured by large and volatile oil revenue flows to the extent that uncertain and volatile flows complicate the management of macroeconomic policies in these countries. Given the exhaustibility of oil reserves, these countries need to address longer-term sustainability and intergenerational equity issues. Conventional fiscal indicators and tools, such as overall and primary balances and Debt Sustainability Analysis (DSA), are not sufficient to make a full assessment of short-term fiscal stances or longer-term fiscal sustainability. The use of more comprehensive fiscal indicators can greatly aid addressing these notable challenges in economies with highly uncertain fiscal revenues such as those in Africa.

The design of sustainability indicators is particularly relevant when it comes to countries which operate in a highly volatile environment. Consider, for example, a temporary appreciation of the real exchange rate. In a country with a large foreign debt, a real appreciation may improve the current fiscal situation by reducing debt servicing in terms of non-tradable goods (i.e. in terms of public sector wages). In contrast, in an economy where the main source of

M. Obinyeluaku (✉)
International Trade Administration Commission of South Africa, Pretoria, South Africa

D.V. Malito et al. (eds.), *The Palgrave Handbook of Indicators in Global Governance*,
https://doi.org/10.1007/978-3-319-62707-6_13

fiscal revenue is a tradable good (such as oil), a real appreciation may well damage fiscal accounts by reducing revenues in terms of non-tradable goods. Hence, in the presence of large temporary fluctuations in the real exchange rate, a reading of the fiscal situation based on current fiscal indicators may lead to a severely distorted assessment of fiscal sustainability. This suggests a need to develop alternative fiscal indicators which may provide a more reliable picture of the underlying sustainability of current fiscal policy, particularly for Africa.

As Africa's budgets and exports are mainly characterised by stochastic revenues, effective management of the continental fiscal revenues is critical for fiscal sustainability and macroeconomic stability.

African countries are heavily dependent on volatile revenues (from aid, oil, exports, and small tax bases) to finance their relatively large total expenditures, making their budgets vulnerable to fiscal shocks. This poses a serious threat both to the sustainability of the continent's budget and to its macroeconomic stability.

"To deal with this volatility challenge, Nigeria, for instance, introduced an oil-price-based fiscal rule in 2004, which was later integrated into … the Fiscal Responsibility Act (FRA) [that] was enacted in 2007" (Ibironke 2013). It was foreseen that this process would smooth government expenditure and stabilise budgetary revenue (Ibironke 2013).

However, a core question in the light of this intended stabilisation function is whether Nigeria's oil-price-based rule can represent an adequate tool to address the macroeconomic conditions that affect fiscal sustainability in Africa. To answer this key question, an analysis of the extent to which the Nigerian oil-price-based fiscal rule has positively influenced the overall fiscal performance and sustainability of public finances in Nigeria post-2004 is essential. This chapter, therefore, evaluates the empirical relevance of the oil-price-based fiscal rule adopted by providing an analysis of post-2004 Nigerian data in order to potentially guide the development of fiscal sustainability indicators in economies with highly uncertain fiscal revenues, such as those in Africa. Analysing the case of Nigeria's fiscal policy promises to be particularly appropriate as Nigeria is the largest country in the continent and it has relied greatly on highly volatile oil revenue since the 1970s.

The chapter is organised as follows. Section 2 "Fiscal Policy with Uncertain Revenues" analyses the implications of introducing fiscal policy rules to control budget dynamics and promote fiscal sustainability in economies with highly uncertain government revenues such as in Africa. Section 3 "Factors Responsible for the Nigerian Volatility Challenge" briefly presents the factors responsible for the challenge of volatility in Nigeria. Section 4 "Nigeria's

Oil-Price-Based Fiscal Rule" gives an overview of Nigeria's oil-price-based fiscal rule. Section 5 "Assessment of the Effectiveness of the Oil-Price-Based Fiscal Rule" assesses the effectiveness of the oil-price rule. Section 6 "Conclusions" concludes the chapter.

Fiscal Policy with Uncertain Revenues

As a monetary policy rule intends to limit the ability of the monetary authority to act discretionally, fiscal policy rules will—if observed—mitigate the government's tendency to abandon previous policy commitments. They seek to confer credibility on the implementation of macroeconomic policies by removing discretionary interventions. Their goal is to achieve trust by guaranteeing that fundamentals will remain predictable and robust regardless of the government in power. Thus, fiscal policy rules are particularly helpful if the government is not able to guarantee a prudent fiscal policy. It therefore seems appropriate to study the sustainability of simple fiscal rules in a case representative of many African countries, where the first source of macroeconomic instability is certainly the dynamics of fiscal policy.

Given the stochastic characteristics of government revenue in Africa, I analyse the implications of introducing fiscal policy rules to control budget dynamics and promote the necessary medium-term budget deficit stability and fiscal sustainability.

One possible way of modelling fiscal policy in Africa is to look at the effect of being dependent on natural resource revenues on the sustainability of two rules—a fixed rule and a variable rule. For example, consider a country in which about 80 per cent of its revenue comes from oil. In this case, we can safely assume that the total gross budgetary revenues for the country are equal to

$$GR_t = P_t\left(\bar{Q}_t\right), \tag{13.1}$$

where GR_t is government revenue, $\bar{Q}_t$ is the quantity of oil, assumed to be fixed,[1] and P_t is its price. Thus, the primary surplus at the end of the budget year is equal to

$$PS_t = P_t\left(\bar{Q}_t\right) - G_t. \tag{13.2}$$

Each year the government has to plan expenditure G_t on the basis of a forecast of oil revenues for the period. If we assume that the price of oil follows a pure random walk, and therefore cannot be predicted, it implies that the best way to look at the predictability of the oil price changes is to assume that oil price equal to $E_t(P_t) = P_{t-1}$. Following this, the expected primary surplus at the beginning of a budget year is

$$E_{t-1}\left(PS_t\right) = E_{t-1}\left(P\bar{Q}_t\right) - E_{t-1}\left(G_t\right) \tag{13.3}$$

Inability to control fiscal revenue introduces a significant element of uncertainty into the budgetary process, equal to the volatility of oil prices v_t. Any fiscal rule in this context should be tested using the budgetary process described by equation (13.3).

Once the government expenditure decision and oil prices are determined, the resulting primary surplus will give the following debt dynamic:

$$D_{t+1} = \left(1 + R_t\right)\left(D_t - PS_t\right), \tag{13.4}$$

where R_t is the real interest rate in period t and D_t is the debt stock at the beginning of the period. Both PS_t and D_t are in real terms. In order to express equation (13.4) in terms of the output ratio, we assume a constant growth rate of output. The path of real output is then given by

$$Y_{t+1} = \left(1 + g_t\right)Y_t, \tag{13.5}$$

where g_t is the constant growth rate. Defining the debt-to-GDP ratio as $d_t = D_t / Y_t$ and combining equations (13.4) and (13.5),

$$d_{t+1}\left[\left(1 + r_t\right) / \left(1 + g_t\right)\right]\left(d_t - ps_t\right), \tag{13.6}$$

where $ps_t = PS_t / Y_t$.

Assuming too, as in Basci et al. (2004), that R_t and g_t have random components, we can define the random variable $r_t + \varepsilon_t$, the growth-adjusted real interest rate, using the following decomposition:

$$1 + r_t + \varepsilon_t = \frac{\left(1 + R_t\right)}{\left(1 + g_t\right)}, \tag{13.7}$$

where r_t is the deterministic component of the real growth-adjusted interest rate and ϵ_t is a zero-mean independently and identically distributed (iid) random variable which represents the interest rate and growth shocks.

Next, we assume that the deterministic component of the growth-adjusted mean real interest rate $r(d_t)$ is an increasing function of the debt-to-GDP ratio (see Cantor and Packer 1996; Hu et al. 2001; Basci et al. 2004):

$$r_t = r\left(d_t\right) \text{with } r'\left(d_t\right) > 0, \tag{13.8}$$

where $r'(d_t)$ represents the first derivative of $r(d_t)$.

Combining (13.6), (13.7), and (13.8), we obtain

$$d_{t+1} = \left(1 + r\left(d_t\right) + \varepsilon_t\right)\left(d_t - ps_t\right), \tag{13.9}$$

where d_t denotes the debt-to-GDP ratio at the beginning of period t, and ps_t denotes the ratio of the primary surplus to GDP in period t. It is assumed that the growth-adjusted mean real interest rate $r(d_t)$ is an increasing function of the debt-to-GDP ratio.

Since the analysis here is limited to a developing country, a linear function of debt stock is assumed for reasons of simplicity[2]:

$$r\left(d_t\right) = \rho d_t \text{ for all } t, \tag{13.10}$$

where $0 < \rho < 1$.

Now, by defining the critical or steady-state debt level (d_c) as

$$E\left[d_{t+1}\right] = d_t = d_c \tag{13.11}$$

and combining (13.9), (13.10), and (13.11), we obtain

$$\rho d_c^2 - \rho d_c ps_t - ps_t = 0. \tag{13.12}$$

Given the dynamic process described by equation (13.9), we need to find the fiscal policy rule that minimises the probability of exceeding the critical debt level in equation (13.12). As in Basci et al. (2004), we consider two

alternative policy rules: a policy rule that stipulates a fixed primary surplus relative to GDP, and one that adjusts the primary surplus required to the level of debt accumulated.

Fixed Fiscal Policy Rule

The fixed primary surplus rule is equal to a constant percentage of GDP, s, in every period: $ps_t = s$ for all t, as

$$ps_t = s = P_{t-1}\left(\bar{Q}_t\right) - G_t. \tag{13.13}$$

By controlling for G_t[3] our fixed expenditure rule now becomes

$$G_t = \left[P_{t-1}\left(\bar{Q}_t\right)\right] - s. \tag{13.14}$$

Equation (13.14) is the level of expenditure necessary to maintain a fixed primary surplus rule.

Variable Fiscal Policy Rule

A variable fiscal rule adjusts the expected level of fiscal surpluses to the outstanding level of debt so that a higher fiscal surplus (a tighter fiscal policy) is set as the debt stock increases. A simple linear expression of this could be

$$ps_t = \sigma d_t \text{ for all } t, \sigma > 0.$$

Substituting σd_t for s in (13.14), our variable expenditure rule will look like

$$G_t = \left[P_{t-1}\left(\bar{Q}_t\right)\right] - \sigma d_t. \tag{13.15}$$

Again, equation (13.15) is the level of expenditure necessary to maintain a variable primary surplus rule.

As will be discussed in the next two sections, the Nigerian oil-price-based rule is a variable fiscal policy rule. According to the former Nigerian Minister of Finance, Ngozi Oknojo-Iweala (2013), this rule is a standard technique commonly used by commodity-dependent countries to protect themselves against the volatility of the oil price.

Factors Responsible for the Nigerian Volatility Challenge

In this section, I discuss some of the main characteristics of the Nigerian economy that have contributed to the country's volatility problem, and which are also shared by many African countries, as identified in the literature.

The Country's Dependence on Uncertain Oil Revenue

Nigeria is heavily dependent on oil revenue to finance over 80 per cent of its total expenditure, making its budget vulnerable to fiscal shocks. This poses a serious threat both to the sustainability of the country's budget and to its macroeconomic stability. Oil windfall induces government spending that is difficult to retrench when the oil revenue falls, distorting government budget allocation pattern, cohesion, and stability, and increase deficits and debt stock that has often created an unfavourable environment for monetary policy (Odularu 2008; Akinlo 2012; Baunsgaard 2003; Obinyeluaku 2008; Ibironke 2013).

Due to a strong fiscal dominance in oil-producing countries, fiscal policy tends to be the main channel for propagating external shocks associated with oil price fluctuations into the non-oil economy. Empirical evidence points to a strong correlation between oil revenue and fiscal expenditure in Nigeria. Obinyeluaku (2014) shows that a higher oil revenue induces higher spending. Some studies show that higher spending exerts pressure on aggregate demand, prices, and the real exchange rate, undermining the non-oil economy (Fasano and Wang 2002). Moreover, oil price volatility transmitted to public expenditure through oil revenue has other undesirable consequences for the non-oil economy:

Macroeconomic Volatility Sharp changes in government spending add to volatility in aggregate demand and prices, abrupt swings in the exchange rate, and increased risks faced by investors in the non-oil sector. Macroeconomic volatility has been shown to have an adverse impact on investment and economic growth (Aizemann and Marion 1993; Gavin 1997). Expenditure volatility associated with fluctuations in oil revenue is found to be a key factor explaining slower growth in oil-producing countries compared to resource-poor countries (Gelb et al. 1988; Auty and Gelb 2001; Bjerkholt 2002).

Expenditure Quality A tendency for the quality of public spending to deteriorate during oil booms has been well documented. Introduction of large-scale

new spending programmes during an oil boom can result in overstretched administrative capacity, a weakening of standards in project selection and evaluation, and even a circumvention of public financial management procedures. The result may be a rapid deterioration in the quality, efficiency, and productivity of public spending. During previous oil booms, some countries undertook ambitious investment projects with low rates of return, politically attractive payoffs, and inadequate screening and execution. Expenditure quality has also been weakened in a number of countries by a proliferation of energy subsidies.

Budget Flexibility Expenditure increases during "good times" tend to benefit politically influential groups (e.g. civil servants, the military, farmers). For example, many oil-producing countries use oil windfalls to increase public sector wages. As these new spending programmes become entrenched, it may become difficult to curtail them when oil revenues drop sharply or dry out. In countries with high levels of statutory outlays, fiscal consolidation is often effected by cutting more productive spending categories, such as infrastructure investment and maintenance, with a possible adverse impact on growth. Another possible budget flexibility concern relates to a weakening of revenue-raising efforts during oil booms, which makes the budget more vulnerable to oil downturns.

The Small Size of the Economy

Nigeria, like any other economy in Africa, is a small open economy. It is a price taker in the world markets. Its domestic interest rate adjusts to that of the world. It has a small gross national income (GNI) per capita (Ibironke 2013).

Basically, small economies like Nigeria usually have difficulties managing external shocks. Unmanaged external shocks bring difficulties and costs to the Nigerian economy.

External imbalances, fiscal and monetary disequilibria, and inflation have been a recurrent problem because expenditure programmes have not been cut when oil prices have fallen. This has been either because the price falls were seen as temporary or because programmes were difficult to stop or reduce at the end of booms. In the 1970s and early 1980s, this problem was so severe that, even before oil prices began to fall, the excess of expenditure over revenue had become persistent, initiating the growth of Nigeria's large stock of external debt. Given that external and internal imbalances cannot be maintained indefinitely, expenditure cuts have been unavoidable. But these cuts have been too late or too costly, or both.

Nigeria fell victim to the spending disease when oil prices and public revenues were high in the 1970s and early 1980s. Its emerging export revenues were spent on the domestic economy, particularly on non-tradable goods, increasing the relative prices of non-tradable goods and wages. Despite favouring the expansion of non-tradable sectors, such as services and construction, this response hurt the development of tradables (other than oil). Thus, Nigeria, a net exporter of agricultural products in the early 1970s, was importing more than US$ 2 billion a year in foodstuffs a decade later.

Private investment also suffered. With the public expenditure programme expanding and contracting at the whim of oil revenues, the volatility and uncertainty that plague oil earnings were channelled to the domestic economy through changes in relative prices and in the associated structure of production. If the oil shock had been permanent, the response would have been the correct one. However, because oil prices are uncertain and highly volatile, investors cannot predict when the next shock will take place. Neither can they predict the direction of the next shock or which sector will be favoured and which one hurt. This uncertainty increases the risk investors face in non-oil activities, reducing the volume of private investment and slowing the growth of the non-oil economy.

There are other macroeconomic costs too. Capital flight is often the private sector's response to a fear that, once oil revenues fall, unsustainable budget deficits will bring inflation and higher future taxes. Moreover, there is often an unproductive political struggle among economic players trying to appropriate windfalls during the booms and to avoid losses during the busts. This process weakened decision-making in Nigeria.

A High Degree of Openness

"There are two sides to openness, namely trade openness… and financial openness..." (Ibironke 2013). Trade openness can be measured as the share of the sum of total exports and imports of merchandise goods and services in gross domestic product (GDP). "The level of Nigeria's ... [trade openness] is relatively high (see, for example, Obinyeluaku 2008). On the other hand, the level of … [financial openness] may be estimated as the ratio of equity-based foreign liabilities to GDP (Calderon et al. 2005)" (Ibironke 2013). The recent shift in policy towards market-oriented systems has led to increasing attention to the development of efficient financial systems in developing countries. The financial sector has a key role in the savings-investment growth race by providing a channel to promote investment by raising and distributing capital. Liberalisation in Nigeria began with the relaxation of entry barriers into the

financial services sector and was followed by a Central Bank relaxation of restrictions on capital inflows/outflows, interest rates, foreign exchange, and bank ownership. Since then, Nigeria has been characterised by trends of increasing liberalisation, greater openness to world trade, and higher levels of financial deepening and integration. This increased openness has motivated increases in private capital inflows and outflows, as is apparent in the fast growth of the country's stock market capitalisation.

A High Degree of Global Integration

Nigeria's economy is highly integrated into the global economy through the process of globalisation. The "country is currently ranked 97th in the overall globalisation rankings of the KOF Swiss Economic Institute" (Ibironke 2013).[4] This high degree of international integration, however, also leads to increased external exposure, as measured by the sensitivity of first and second moments of economic growth to openness and foreign shocks. This vulnerability is particularly important in Nigeria due to its production specialisation, non-diversified income sources, unstable policies, incomplete financial markets, and weak institutions.

The Emerging Market Feature

Nigeria has been described by the international financial institutions (the International Monetary Fund, IMF, and the World Bank) as one of the 11 countries to watch out for in the next decade (IMF 2013a, b). Lucrative investment ventures are rife, and their development potential continues to rise.

Impressive growth has been recorded in the emerging capital markets in this petrodollar-rich sub-Saharan nation of prestigious natural resources, offering very attractive opportunities for market operators and investors. Regarded as the second most impressive after South Africa, these markets, which had hitherto been closed to foreign investors (functioning solely as a government auction/trading post for treasury securities and equity shares of statutory corporations and foreign subsidiary companies), were restructured to allow for the participation of free market institutions after 1999 when the country returned to civilian rule. This was the best time to invest in the country, especially in the financial services industry as an engine of wealth creation. Several foreign companies lucratively partnered with indigenous brokers to trade in equities and government bonds. The problems that occurred in the

advanced capital markets in the world—USA, Europe, and Asia—have generated compelling arguments for the participation of foreign investors in emerging capital markets. These markets offer profit-making opportunities for asset diversification. The market potential in the country is huge, with investment opportunities even in the real sector—power, housing, agriculture, transportation, and tourism.

Nigeria's relevance in the oil market has brought it much attention. It is the 5th largest exporter of crude oil to the USA. In addition, the country's GDP has grown at an average annual rate of 8 per cent for five consecutive years. Foreign investment has poured in, especially from the USA, which is Nigeria's largest foreign investor. However, the bulk of investment is in the oil sector of the economy.

Nigeria's Oil-Price-Based Fiscal Rule

Nigeria is heavily dependent on oil revenue to finance over 80 per cent of its total expenditure, making its budget vulnerable to fiscal shocks. A strong deficit and debt bias stemming from government revenue volatility hence poses a serious threat to the country's budgetary sustainability. An oil windfall induces government spending, which in turn is difficult to reduce when the oil flow declines, distorting government budget allocation patterns and increasing deficits and debt.

To deal with this challenge of volatility, Nigeria, for instance, introduced an oil-price-based fiscal rule in 2004, which was later integrated into the Fiscal Responsibility Act (FRA) that was enacted in 2007 (Nigerian National Assembly 2007). It was expected that this process would smooth government expenditure and stabilise budgetary revenue (Ibironke 2013).

Through the determination of a domestic crude oil benchmark for budgeting, various endogenous and exogenous factors became influential in pegging the benchmark in the Nigerian budgeting process. These factors include the social and economic objectives of the government, the costs of oil production, joint-venture agreement considerations, oil production during the contract period, non-oil sector viability, and the overall fiscal stance of the government. The interplay of these variables affects not only the crude oil benchmark for budgeting, but also the government's revenue stream projections in its fiscal planning.

Put differently, the Nigerian oil-price-based fiscal rule stresses "that annual fiscal expenditure is restrained through a reference oil price" (Ibironke 2013). It uses projections of the oil price that are lower than the expected

international price over the budget period. The benchmark price is a result of some rigorous analysis that develops a ten-year moving average oil price (Ibironke 2013). In 2004, based on these calculations, the government budgeted at a price of US$25. Subsequently, the figure was US$30 in 2005, US$35 in 2006, US$40 in 2007, US$72 in 2012, and US$75 in 2013.

When oil revenues are high such that the actual oil price is above the benchmark price, the resulting surplus "is kept in a special 'Excess Crude Oil Account' (ECA), which is also known as the 'Sovereign Wealth Fund'" (Ibironke 2013). When oil revenues are low, the ECA would finance the shortfall (Ibironke 2013). According to the previous administration's Minister of Finance, Ngozi Oknojo-Iweala (2013), this way of proceeding is a standard technique commonly used by commodity-dependent countries to protect themselves against oil price volatility.

The benchmark oil price is usually determined by the executive arm of the government. This method, and especially the role of the government within it, has, however, been criticised for being highly subjective, lacking transparency and not being based on generally accepted accounting principles (GAAP). This has often resulted in heated debates and intensive negotiations between the executive and the legislative arms of government, and frequently led to delays in the release of the appropriation bill. For example, while legislative actors had proposed a US$79 per barrel oil price benchmark for the 2014 budget, the executive branch required US$75, but ended up with $77.5 in the final appropriation. The 2014 budget was delayed due to this mismatch.

Following consultations with various stakeholder groups, including governors and the National Assembly, the president finally approves the benchmark budgeted oil price, which in turn places a limit on government expenditure.

During the budget process, ministries, departments, and agencies (MDAs) of government receive "expenditure envelopes" from which they are to satisfy their financial needs, including salaries. These expenditure envelopes cater for the priority level accorded to the services to be delivered by the MDAs as articulated in their Medium-Term Sector Strategies (MTSS) against the background of the priorities of the federal government (FG) as documented in the National Economic Empowerment and Development Strategy (NEEDS) and the Sustainable Development Goals (SDGs). The allocation of these envelopes is determined by the benchmark oil-price-based rule.

In general, government oil revenues are allocated to four main areas: (a) federal, state, and local budgets, and extra-budgetary funds; (b) cash calls from the Nigerian National Petroleum Corporation, NNPC (to finance expenditure and investment in the oil sector); (c) the fuel subsidy; and (d) the Excess Crude Account (ECA). Allocations to the ECA can be either positive

(accumulation) or negative (drawdowns), which augment the oil revenue allocated to other directions. The dynamics of the distribution of oil revenues reflect different factors and policies that form priorities in the year in question. For example:

In 2008, oil revenues surged due to record high prices, supporting exceptionally large collections to budgets and an accumulation to the ECA (3.4 per cent of GDP).
In 2009, in the light of a sharp decline in output and revenues, significant allocations from the Federation Account and fuel subsidy payments were financed by large drawdowns of the ECA (5.7 per cent of GDP).
In 2010, an increase in oil prices was not sufficiently significant to finance large increases in allocations to budgets and higher fuel subsidy payments, entailing another drawdown of the ECA of 2.7 per cent of GDP.
In 2011, oil prices and revenues increased notably. Large increases in allocations to budgets and a major surge in payments of the fuel subsidy (4.6 per cent of GDP) limited accumulation in the ECA to only 1.3 per cent of GDP.
In 2012, oil output and revenues dropped but the government did not deplete the ECA, which accumulated another 1.5 per cent of GDP.

A thorough analysis of the ECA from its inception in 2004 onwards shows that the FG has continuously augmented its distributable revenue from the ECA due to shortfalls in the production of petroleum products and tax income. Between 2005 and 2008, the savings in the ECA rose from $5.1 billion to $20 billion, but due to continual drawdowns the account cascaded down from US$20 billion to US$4.1 billion in 2014.

Assessment of the Effectiveness of the Oil-Price-Based Fiscal Rule

Whether an oil-price-based fiscal rule such as the Nigerian one can serve to address the macroeconomic conditions that affect fiscal sustainability in Africa also depends on the extent to which this rule is assessed to have positively influenced the macroeconomic performance of Nigeria after 2004. The following section looks at the post-2004 performance of the Nigerian oil-price-based fiscal rule with respect to its original objectives and its impact on fiscal outcomes. The analysis is based on (a) the extent of macro-fiscal vulnerabilities, (b) controllability and fiscal consolidation, (c) the cyclical properties of fiscal policy, (d) the balance of payments, and (e) inflation.

The Extent of Macro-Fiscal Vulnerabilities

Does Nigeria face a volatile macroeconomic environment? And, if so, is this environment more volatile than before the introduction of the oil-price-based fiscal rule in 2004? Tables 13.1 and 13.2 summarise differences in variability by listing the volatilities of government expenditure and revenue together with other macro variables, including annual consumer price inflation, the nominal effective exchange rate and GDP growth.

As one of the countries with the highest revenue volatilities, a key priority for fiscal policy should be to protect the budget from such volatility. In particular, the oil-price-based fiscal rule is expected to restrain government expenditure through oil revenue smoothing, which involves setting a volatility-absorbing reference oil price through which resource revenues will be channelled into the budget in order to avoid the destabilisation effect of fiscal shocks on monetary policy and macroeconomic stability. In practice, Nigeria has, on average, not only been able to limit volatility in expenditure since 2004 but revenue volatility has even exceeded expenditure volatility (see Table 13.1). This suggests that the oil-price rule is capable of limiting the impact of the volatility of revenue on the budget.

Furthermore, Table 13.2 shows that on average the volatilities of inflation, the nominal effective exchange rate, and output are lower after 2004

Table 13.1 Volatility of government revenue and expenditure, 1995–2014 (percentage of GDP)

	Standard deviation		Coefficient of variation	
	Pre-2004	Post-2004	Pre-2004	Post-2004
Revenue	2.92	8.13	21.7	49.5
Expenditure	5.44	2.23	33.1	16.5

Source: Author's estimation based on IFS
Note: Pre-2004 = 1995–2003 and Post-2004 = 2005–2014, excluding 2009
The coefficient of variation is measured as standard deviation/mean

Table 13.2 Volatility of other macro variables

	Coefficient of variation		Averages	
	Pre-2004	Post-2004	Pre-2004	Post-2004
Inflation rate	87.90	34.48	18.5	10.7
Exchange rate	63.04	11.41	53.7	60.9
GDP growth	49.49	26.12	3.1	6.1

Source: Author's estimation based on IFS and Central Bank of Nigeria (CBN)
Note: Pre-2004 = 1995–2003 and Post-2004 = 2005–2014, excluding 2009

compared to the pre-2004 period, with the country almost doubling its GDP growth post-2004. As a result, Nigeria seems to have been facing a less volatile macroeconomic environment since the introduction of the oil-price rule in 2004.

Controllability and Fiscal Consolidation

A strong deficit and debt bias stemming from government revenue volatility poses a serious threat both to the sustainability of Nigerian fiscal policy and to macroeconomic stability. The oil-price rule is considered an important element of budgetary consolidation as it limits the political scope and reduces the deficit and debt bias while ensuring that a fiscal reserve of adequate size is accumulated to protect Nigeria from oil price volatility and prevent the government from losing control over its budget.

The evidence shown in the previous section suggests that low expenditure volatility seems to have coincided with better fiscal outcomes in Nigeria after the introduction of the oil-price rule. The post-2004 period has witnessed a persistent decline in fiscal deficit in Nigeria (Fig. 13.1).[5] Public revenue consistently exceeded expenditure except for 2010, 2013, and 2014. Government revenue increased from an average of 13.5 per cent of GDP before the introduction of the oil-price rule (pre-2004) to 16.4 per cent after 2004. At the same time, public expenditure declined to reach 13.5 per cent of GDP after 2004. Consequently, the fiscal deficit turned into a surplus and debt stock to GDP declined significantly over the same period under review (Fig. 13.1).

According to the IMF's Debt Sustainability Analysis, the risk of debt distress remains low (IMF 2015). Total annual debt post-2004 (pre-2004) was 13.5 per cent (48.8 per cent) of GDP, with external debt, mostly from International Financial Institutions (IFIs) on concessional terms, at only 3.7 per cent (34.8 per cent) of GDP. However, about 30–40 per cent of the FG domestic debt was held by non-residents at the end of 2013. Moreover, FG debt servicing on total public debt was 5.6 per cent (9.6 per cent) of general government revenue.

However, a vital question remains whether the adjustments occurred on the right side of the budget. This is of importance as the composition of the adjustment affects the success and durability of government budget consolidation (Von Hagen et al. 2002; Perotti et al. 1998). Von Hagen et al. (2002) and Perotti et al. (1998) have demonstrated that consolidation relying on expenditure cuts is more likely to lead to a permanent reduction in deficit

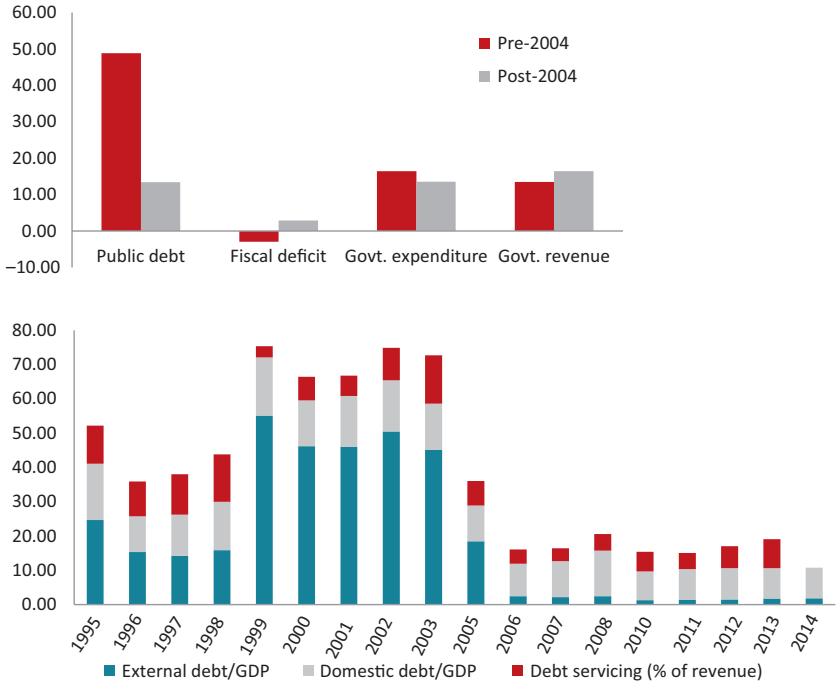

Fig. 13.1 Nigerian fiscal indicators, 1995–2014 (percentage of GDP, except for debt servicing). Source: Author's estimation based on data from the Central Bank of Nigeria (CBN)

Table 13.3 Contribution to fiscal surplus after 2004

	Pre-2004	Post-2004	Contr. to surplus in post-2004
Fiscal deficit	**–2.93**	**2.91**	
Expenditure	**16.42**	**13.53**	**–2.89**
Capital	7.56	3.66	–3.89
Recurrent	8.86	9.86	1.01
Revenue	**13.49**	**16.44**	**2.95**
Oil	9.99	13.25	3.26
Non-oil	3.50	3.19	–0.31

Source: Author's estimation based on data from the Central Bank of Nigeria (CBN)

and ensure fiscal sustainability than consolidation based on raising additional revenue. This finding calls for disaggregation of the main deficit components in order to ascertain what is driving the adjustments in the post-2004 period. The results are presented in Table 13.3.

This table contains three clear messages. First, the size of the fiscal surplus (2.91 per cent) witnessed after the introduction of the oil-price rule (post-2004) is virtually driven by both expenditure cuts (−2.89 per cent) and

increased revenue (2.95 per cent). The second message concerns the composition of the expenditure cuts. Cuts in capital expenditure contribute more than 100 per cent to the decline in expenditure after 2004. Recurrent expenditures, by contrast, behave in the opposite way, increasing overall expenditure by 1 per cent. The third point concerns disaggregation of revenue. The contribution of revenue to the fiscal surplus after 2004 is driven by oil revenues. However, non-oil revenues declined by 0.31 per cent over the same period. With oil production below the budget target of 2.38 million barrels per day (bpd) and the continual decline in oil prices, Nigeria needs to reduce its dependency on oil revenues and ensure there are measurable increases in non-oil revenues, and also reverse the trend of increasing recurrent expenditure and declining capital expenditure.

The Cyclical Properties of Fiscal Policy

A high dependence on resource exports is often associated with lower growth and greater economic instability due to "boom-bust" government spending resulting from highly volatile commodity prices. Countercyclical fiscal policy (expansionary when growth is below the trend and contractionary in good times) is generally desirable because it helps to smooth output volatility.

By introducing automatic revenue stabilisers, Nigeria's oil-price rule and the resultant ECA are expected to ensure a more countercyclical fiscal policy. Positive allocations to the ECA (or accumulation) are likely to mitigate the overheating of the country's economy from exceptionally high oil prices during "boom" periods, while negative allocation to the ECA (or drawdowns) can ensure macroeconomic stability during "bust" periods by maintaining strong growth in domestic demand and GDP.

The extent to which policies are countercyclical is typically measured by correlations between cyclically adjusted measures of government activity and the output gap.[6] The output gap is measured as deviations between output levels from their long-run trends using the Hodrick-Prescott filter (Kaminsky et al. 2004). Economic downturns or recessions are defined as periods where output gaps are negative (or where growth is below the trend). The opposite is the case for an economic boom. Table 13.4 summarises the resulting cyclical fiscal patterns.

Figure 13.2 indicates that fiscal policy in Nigeria remained mostly procyclical even after the introduction of the oil-price rule. Through the establishment of the ECA fiscal reserve, the country made a positive step during 2005–2008 that successfully insulated it from the sharp swings in oil prices

Table 13.4 Cyclical fiscal policy patterns

		Real spending growth	
		Positive	Negative
		(Fiscal expansion)	(Fiscal contraction)
Output gap	**Negative**	Countercyclical	Procyclical
	Positive	Procyclical	Countercyclical

Source: Author's own design

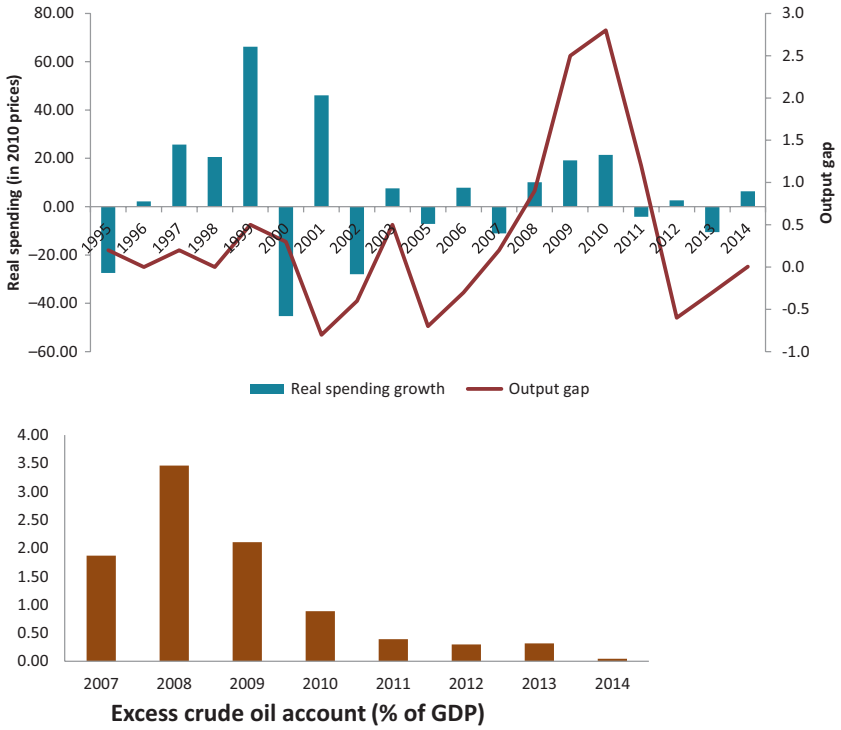

Fig. 13.2 Cyclical properties of fiscal policy. Source: Author's estimation based on data from the Central Bank of Nigeria (CBN)

during this period. However, despite the recovery in oil prices in 2010, Nigeria expanded its fiscal stimulus significantly, increasing consolidated spending by 21 per cent in real terms and drew down the ECA at the same time that many other exporters were building back their reserves.

The limited degree of accumulation undermined the ability of the country to ensure stability during the period of declining oil prices and output in 2013–2014. For example, the ECA reached $2.0 billion by the end of 2014—well below the $6.3 billion required to cover a one-half standard deviation shock to oil receipts.[7]

The Balance of Payments

Nigeria's high dependence on inherently volatile oil revenues presents major balance of payments risks to the country. A sharp decline in oil prices not only has a strong impact on its current account but on the capital account too, as the general attitude of investors towards the country critically depends on oil prices and the capacity to manage the risks of oil-price volatility.

Figure 13.3 shows that Nigeria's balance of payments position has strengthened along with the improved management of fiscal policy since the introduction of the oil-price rule in 2004. Between 2005 and 2008, the balance of

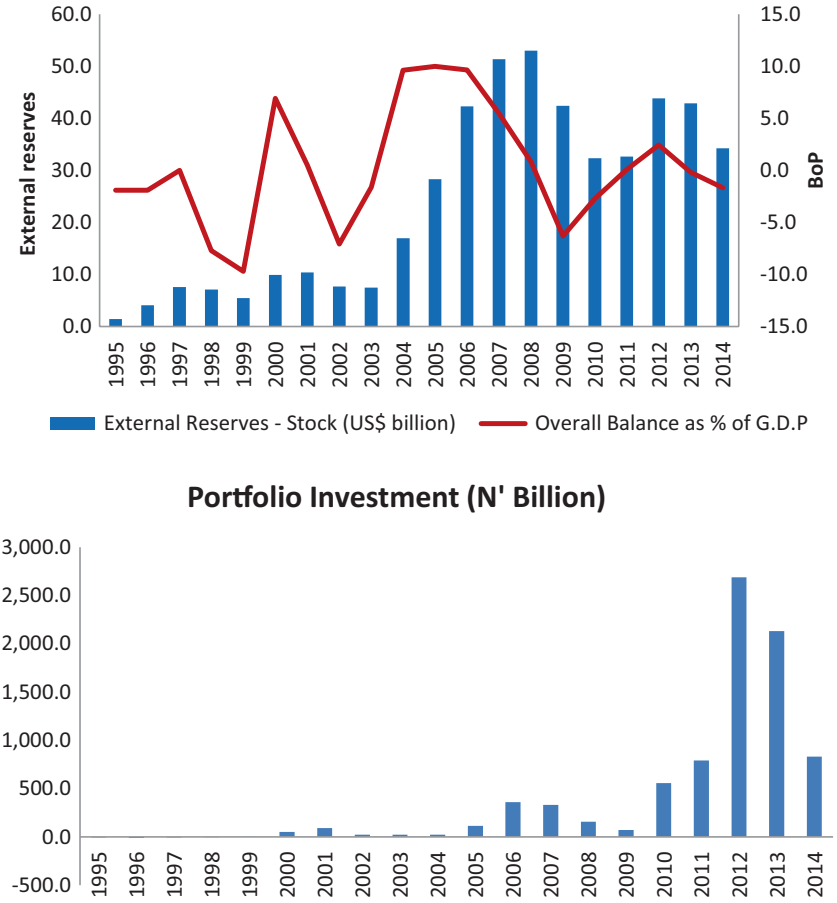

Fig. 13.3 Nigeria's BOP and external reserves, 1995–2014. Source: Author's estimation based on data from the Central Bank of Nigeria (CBN)

payments was in surplus, allowing the Central Bank to build its foreign reserve position from US$ 6.8 billion on average in 1995–2003 to US$ 43.7 billion in 2005–2008.

However, despite the recovery in oil prices in 2010–2011, which enabled many other oil-dependent emerging markets to restore a balance of payments equilibrium, Nigeria's balance of payments remained in deficit and the country lost US$ 0.3 billion in foreign reserves. As previously discussed, with stronger oil prices, Nigeria actually expanded its fiscal stimulus in 2010–2011, drawing not only on higher oil revenues but also on the balance of its ECA. Consequently, imports recovered and the balance of payments declined, putting pressure on Nigeria's currency, the Naira.

In 2012, the country began to re-accumulate its fiscal reserves, which had a notably positive effect on the expectations of investors. Combined with other factors, this development started attracting substantial foreign inflows to the government bond market. Portfolio investment inflows to Nigeria increased from 792.4 billion Naira in 2011 to 2.7 trillion in 2012. These inflows, against the backdrop of tighter fiscal policy, primarily explain the widening of the country's balance of payments surplus in 2012, despite somewhat weaker oil prices.

By 2013 and 2014, the continued decline in oil prices and revenues had led to a limited degree of accumulation, and drawdowns, adversely affecting investor confidence, culminated in an abrupt reverse of the short-term inflows, thereby magnifying the oil-price-related balance of payments swings.

Nigeria faces a medium-term challenge in managing its balance of payments. Given the present declining oil prices and export demand, the pace of import growth is likely to exceed export growth for some years ahead. Thus, the prevailing balance of payments deficit is very likely to continue, and more exchange rate flexibility might be necessary over the longer term. Moreover, its economy needs to be diversified.

Inflation Performance and Monetary Policy

Consumer price inflation (CPI) has remained stubbornly high in Nigeria (Fig. 13.4). Contrary to some expectations, given the tightening of monetary policy, CPI (year-on-year) has been at an average of 11 per cent during the post-2004 period compared to 20 per cent between 1995 and 2003.

In a context of poor weather conditions in Nigeria and increases in world food prices, high food prices drove inflation up in 2008. Despite declining food and commodity prices, the continued high inflation in Nigeria in 2010 no doubt reflected the strong fiscal expansion in that year. Monetary policy

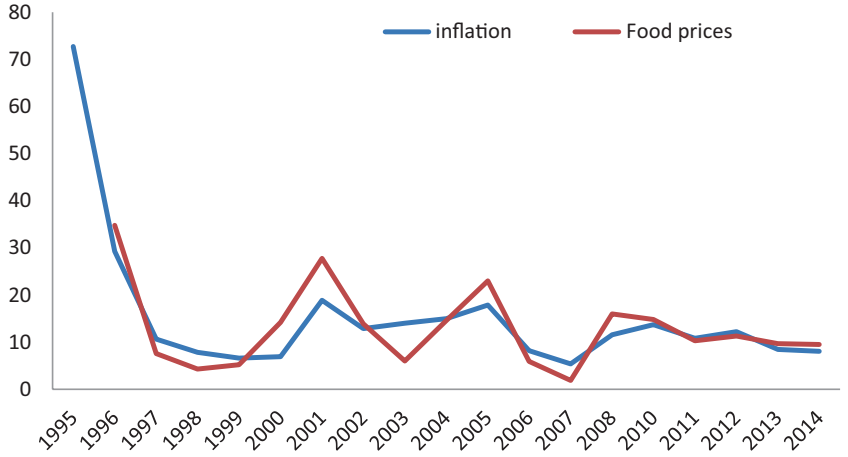

Fig. 13.4 CPI and food price inflation in Nigeria (year-on-year). Source: Author's estimation based on data from the Central Bank of Nigeria (CBN)

was also eased in the context of the Nigerian banking crisis that unfolded in 2009, but without any corresponding rapid expansion in money supply or credit that could have been inflationary. The inflation rate dropped in 2011 in the context of both fiscal and monetary tightening, but increased again in 2012. Part of the explanation for this development concerns one-off effects on inflation of administrative increases in petrol prices (50 per cent reduction of the fuel subsidy) and electricity tariffs. In addition, severe flooding and security challenges in parts of the country reduced the supply and trading of some goods. Driven by the decline in food and commodity prices, the inflation rate began a steady fall again in 2013.

Conclusions

This chapter has analysed whether the Nigerian oil-price-based fiscal rule is able to adequately address the macroeconomic conditions that affect fiscal sustainability in economies with highly uncertain fiscal revenues, such as those in Africa. The findings show that it has been effective in Nigeria and that it therefore can assist in contributing to improved macroeconomic conditions in other similar economies in Africa. The rule introduces some element of flexibility in the way expenditure is planned, and offers a much less stringent constraint on the policy-maker. The high degree of macroeconomic stability in Nigeria in the past decade directly reflects important progress in this direction, but remaining institutional weaknesses still need to be addressed.

The country needs to overcome the political bias and private interests involved in the determination of the benchmark oil price. The government of Nigeria has been finding it increasingly difficult to adhere strictly to and accurately forecast revenue accruals based on the benchmark oil-price rule for each fiscal year. It sets a benchmark price that either overshoots or undershoots the level that is consistent with the expected revenue for the fiscal year or 3-year medium-term expenditure framework. Since the success of this fiscal rule is dependent on the consistency with which the benchmark mimics the volatile and exogenously determined international price of crude oil, an accurate forecast of this crude oil price is crucial to the continued achievement of the predetermined overall fiscal performance in Nigeria. It would be constructive to limit the yearly debate and conflicts surrounding budget preparation over the choice of an appropriate benchmark price through legislation that fixes the allocation rule for a longer period of time.

At present, Nigeria faces both the challenge of (highly likely) declining oil revenues relative to GDP and the imperative to build a sufficient fiscal reserve to ensure macroeconomic stability. Planning a high rate of real growth in the distribution of oil revenue to budgets and the resultant depletion of the oil savings account (ECA) would put fiscal sustainability at risk. As such, the recent withdrawal of the fuel subsidy is a step in the right direction. A sufficient reserve has to be accumulated to insulate the country from sharp swings in oil prices. International experience in oil-dependent countries suggests that countercyclical fiscal policy is the key to conquering the "oil curse" of periodic instability.

Having said that, non-oil revenue is just 4.5 per cent of non-oil GDP—compared to an average of 10–15 per cent of non-oil GDP for other oil producers. Given that oil revenues are due to become increasingly low relative to the size of the Nigerian economy, non-oil revenue mobilisation will become a key fiscal priority in the period ahead. The task of building a strong domestic tax system at the federal and subnational levels and efforts to diversify the economy become increasingly critical.

With the above lessons in mind, the Nigerian oil-price-based fiscal rule can function as a pan-African indicator of fiscal sustainability. It can play a role in stabilising expenditure programmes at levels consistent with the necessary medium-term deficit stability. The oil-price-based fiscal rule will, if observed, mitigate the tendency of democratic governments to abandon previous policy commitments as it introduces a long-term horizon to governments' often short-sighted decision-making processes. It seeks to confer credibility on the conduct of macroeconomic policies by removing discretionary interventions. The goal is to achieve trust by guaranteeing that fundamentals will remain predictable and robust regardless of the government in power.

Notes

1. As it is exogenous and determined by OPEC not the government.
2. It is also assumed that the real interest rate is independent of the fiscal rule adopted.
3. We cannot control for P_t (Q_t) due to oil price volatility.
4. The overall globalisation index comprises economic, social, and political globalisation. Nigeria ranked 82nd, 184th, and 25th, respectively, according to the 2015 study.
5. The fiscal deficit consists of the primary deficit and interest payments on outstanding government debt.
6. Measured as the difference between actual and potential growth.
7. However, the fiscal authorities responded swiftly to the oil-price developments by submitting a revised 2015 Medium-Term Expenditure Framework (MTEF) in December, with a benchmark oil price of $65 per barrel (pb) compared to the $78 pb in the original MTEF (submitted in October).

References

Aizemann, J., & Marion, N. P. (1993). Policy Uncertainty, Persistence, and Growth. *Review of International Economics, 1*(2), 145–163.

Akinlo, A. E. (2012). How Important Is Oil in Nigeria's Economic Growth? *Journal of Sustainable Development, 5*(4), 165–179.

Auty, R. M., & Gelb, A. H. (2001). Political Economy of Resource-Abundant States. In R. M. Auty (Ed.), *Resource Abundance and Economic Development. Study Prepared for the World Institute for Development Economics Research of the United Nations University* (pp. 126–146). New York: Oxford University Press.

Basci, E., Fatih Ekinci, M., & Yulek M. (2004). *On Fixed and Variable Fiscal Surplus Rules* (IMF Working Paper, WP04/117). Washington, DC: International Monetary Fund. Retrieved from https://www.imf.org/external/pubs/cat/longres.aspx?sk=17489.0

Baunsgaard, T. (2003). *Fiscal Policy in Nigeria: Any Role for Rules?* (IMF Working Papers, WP/03/155). Washington, DC: International Monetary Fund. Retrieved from https://www.imf.org/external/pubs/cat/longres.aspx?sk=16728.0

Bjerkholt, O. (2002). *Fiscal Rule Suggestions for Economies with Non-Renewable Resources.* Paper Prepared for the Conference "Rules-Based Fiscal Policy in Emerging Market Economies", 14–16 February. Oaxaca.

Calderon, C., Loayza, N., & Schmidt-Hebbel, K. (2005). *Does Openness Imply Greater Exposure?* (World Bank Policy Research Working Paper WP/3733). Washington, DC: World Bank.

Cantor, R., & Packer, F. (1996). Determinants and Impact of Sovereign Credit Ratings. *Federal Reserve Bank of New York Economic Policy Review, 2*(2), 37–54.

Fasano, U., & Wang, Q. (2002). *Testing the Relationship Between Government Spending and Revenue: Evidence from GCC Countries* (IMF Working Paper 02/201). Washington, DC: International Monetary Fund.

Gavin, M. (1997). *A Decade of Reform in Latin America: Has It Delivered Lower Volatility?* (Inter-American Development Bank Working Paper Green Series, No. 349). Washington, DC: Inter-American Development Bank.

Gelb, A., et al. (1988). *Oil Windfalls: Blessing or Curse?* New York: Oxford University Press for the World Bank.

Hu, Y. T., Kiesel, R., & Perraudin, W. (2001). The Estimation of Transition Matrices for Sovereign Credit Ratings. *Journal of Banking and Finance, 26*(7), 1353–1406.

Ibironke, A. (2013). *How Effective Is the Nigerian Oil-Price-Based Fiscal Rule?* (Research Paper). Newcastle: Newcastle University Business School. Retrieved from https://editorialexpress.com/cgi-bin/conference/download.cgi?db_name=CSAE2014&paper_id=888

International Monetary Fund (IMF). (2013a). *International Financial Statistics*. Washington, DC: IMF. Retrieved from http://www.imf.org/en/Data

International Monetary Fund (IMF). (2013b). *World Economic Outlook*. Washington, DC: IMF. Retrieved from www.imf.org

International Monetary Fund (IMF). (2015, February). *Staff Report for the 2014 Article IV Consultation-Debt Sustainability Analysis*. Washington, DC: IMF.

Kaminsky, G. L., Reinhart, C. M., & Végh, C. A. (2004). *When It Rains, It Pours: Procyclical Capital Flows and Macroeconomic Policies* (National Bureau of Economic Research, NBER, Working Paper No. 10780). Washington, DC: NBER. Retrieved from http://www.nber.org/papers/w10780

Nigerian National Assembly. (2007, July 19). *Fiscal Responsibility Act, 2007: Act No. 31*. Abuja: Nigerian National Assembly.

Obinyeluaku, M. (2008). *Fiscal Policy Rules for Managing Oil Revenues in Nigeria.* Paper Presented at the CSAE Conference on Economic Development in Africa, 16–18 March. Oxford: St Catherine's College, Oxford University.

Obinyeluaku, M. (2014). Monitoring Fiscal Sustainability in Africa. In A. B. Elhiraika (Ed.), *Regional Integration and Policy Challenges in Africa* (pp. 278–301). Basingstoke: Palgrave Macmillan.

Odularu, G. O. (2008). Crude Oil and the Nigerian Economic Performance. *Oil and Gas Business Journal*, (1). Retrieved from http://ogbus.com/article/crude-oil-and-the-nigerian-economic-performance/

Okonjo-Iweala, N. (2013). Ngozi Okonjo-Iweala Justifies Nigeria's Budget Oil Price. *African Spotlight*. Retrieved from www.africanspotlight.com/2012/10/15/ngozi-okonjo-iweala-justifies-nigerias-2013-budgeted-oil-price/

Perotti, R., Strauch, R., & Von Hagen, J. (1998). *Sustainability of Public Finances* (No. 1781). Centre for Economic Policy Research.

Von Hagen, J., Hughes Hallet, A., & Strauch, R. R. (2002). Budgetary Consolidation in Europe: Quality Economic Conditions and Persistence. *Journal of the Japanese and International Economies, 16*(14), 512–535.

14

Moving Beyond Traditional Indicators of Fiscal Sustainability: Examples from Locally Chosen Indicators

Daphne T. Greenwood

Introduction

This chapter first reviews widely used definitions of fiscal health or soundness, including the well-known ratio of national debt to national income (debt/GNI). It then examines the meaning of "fiscal sustainability" in the larger context of sustainable development, which was defined by the United Nations' Brundtland Commission (1987) as producing and consuming in ways that will allow future generations to have at least the same standard of living as exists today. The future orientation of sustainable development requires that attention be paid to all the capital stocks which produce well-being (Goodwin 2003). This includes more than investment or depletion of private manufactured capital, both of which are in the national income accounts. The third section of the chapter explains these other capital stocks—public infrastructure, natural resource capital, human capital, and social capital—in more detail as well as how fiscal sustainability depends on their health Greenwood (2010). The fourth section discusses the types of indicators that are consistent with sustainable development theory, with examples taken from locally based indicator sets around the world. Some of these measures, such as the poverty rate or housing affordability, are measured in financial terms. Others, such as the energy efficiency of public buildings or the percentage of land surface impervious to water are not monetised in local indicator sets, as they are in the Inclusive Index of Well-Being (IWI).[1]

D.T. Greenwood (✉)
University of Colorado at Colorado Springs, Colorado Springs, CO, USA

D.V. Malito et al. (eds.), *The Palgrave Handbook of Indicators in Global Governance*,
https://doi.org/10.1007/978-3-319-62707-6_14

However, taken as a whole, these indicators may be more accurate in predicting fiscal sustainability than financial indicators such as debt/GNI have been, because they directly affect the future economic capacity and expenditures, which will affect the need to borrow. The last section of the chapter explains how local indicators of capital stock size and quality are chosen, collected, and used by local governments and how national governments interested in fiscal sustainability could build on this by expanding the measures they use beyond debt/GNI. First, let us turn to the specific definitions of fiscal soundness used by various governments.

Fiscal Soundness: Definitions and Theory

The ratio of national public debt to national income (debt/GNI) is often used to measure the ability to service debt.[2] A growing ratio, "especially if the level of debt is already high, may suggest that a country is on an unsustainable path", according to the Policy Development and Review Department of the International Monetary Fund (IMF 2000).[3] However, many other publications from the IMF and United Nations (UN) caution against using a single measure like debt/GNI as a reliable indicator of fiscal sustainability, as reviewed in other chapters of this handbook. The debt/exports or debt/government revenue ratios are sometimes used as alternatives to debt/GNI, but have many of the same problems. Below, several other definitions of fiscal soundness are explored, beginning with the most restrictive and proceeding to those which are most compatible with sustainable development.

Definitions

According to the European Commission (EC) Report on Fiscal Sustainability (2012), "sustainability of fiscal policies is the ability to continue now and in the future, current policies without change regarding public services and taxation and without causing the debt to rise continuously as a share of GDP" (p. 17). There are two major problems with this very restrictive definition. First, while appropriate for a sub-national government without stabilisation responsibilities (Greenwood and Holt 2015, pp. 118–124), it can be used to support cuts in spending or increases in taxes at a time when countercyclical public spending is needed. This was the source of austerity policies in the European Union (EU) countries during the Great Recession. In the EU countries, spending cuts and tax increases made to prevent the numerator (public debt)

from rising worsened the problem by causing declines in the denominator (national income). A second problem with the definition is its focus on "public service expenditure" and inattention to spending that represents "investment in public capital". It appears to treat all spending equally without accounting for either (a) stabilisation expenditures during recessions or (b) necessary investments in public capital.

The International Public Sector Accounting Standards Board (IPSASB), of which the European Commission is a member, has a different definition. Fiscal sustainability is "the ability of an entity to meet service delivery and fiscal commitments both now and in the future" (IPSASB 2011, p. 5). This does not establish a constraint on changing services or taxation or of maintaining a particular debt/GNI ratio. Instead, the IPSASB definition points to using a wide range of data to determine appropriate fiscal action under various conditions, such as financial and non-financial information on economic and demographic conditions, and expected changes in age, longevity, gender, income, educational attainment, and morbidity (IPSASB 2011, p. 6).

In the US, the Congressional Budget Office (CBO) describes a long-term structural budget imbalance between an existing tax system and projected expenditures as "fiscal stress". This could reflect unwillingness to change the tax structure (or make budget cuts) despite adequate capacity. The CBO (2010) uses "fiscal health" to indicate whether there is sufficient underlying capacity to repay additional debt. This capacity comes from capital stocks and the level of potential national income.

The Government Accounting Standards Board (GASB) in the US goes even further in its definition for state and local governments, and includes "ability and willingness to generate inflows of resources necessary to honor current service commitments and to meet financial obligations as they come due, without transferring financial obligations to future periods that do not result in commensurate benefits" (GASB 2008, pp. 5–6). This parallels the CBO's emphasis on capacity and willingness.[4] The GASB criteria distinguish debt that finances capital from debt that finances operating expenses. Another term from the sub-national level, "fiscal slack", refers to the share of budget spent on capital expenditure (Hendrick 2004), which provides a way to make short-term cuts in a fiscal crisis without affecting operating budgets.

Theoretical Grounding for Fiscal Sustainability Measures

It is clear that commonly used operational definitions of fiscal soundness vary. But they also tend to lack grounding in economic theory. Most analytical

discussions of fiscal sustainability are based on models that only reference the stock of government debt, the interest rate on that debt, and projected primary surpluses and deficits (Chalk and Hemming 2000, pp. 62–64). If deficits exceed surpluses over some (unspecified) period of time, the model shows a higher probability that investors will demand a higher rate of interest on government bonds at some point—and perhaps even refuse to purchase them.

The problem is that most models analyse deficits, surpluses, and interest rates outside of the full context of the economic and political system. In addition, since few national governments separate capital and operating budgets, the models fail to consider what is being financed by new debt. Investments in human, physical, or natural infrastructure that will bring higher productivity, income, and tax revenue in the future (or even lower public expenditure, as will be explored later in this chapter) often require issuing new that raises debt/GNI above its typical level. However, the effects on fiscal sustainability are far different when compared with spending on goods and services that did nothing to raise the capital stocks that will be the basis of future national income.

In addition, national governments are responsible for macroeconomic stabilisation. Temporary deficits needed to underwrite stimulus spending in recessions will cause public debt to increase in the short run—at the same time that GNI is falling. But as GNI grows in response to the new spending, the ratio of debt to GNI will change rapidly as the denominator grows. The long-term ratio is likely to be far different from that in the short term, with the possibility of eventual surpluses that could be used to retire debt.

In short, relying on purely financial indicators of fiscal soundness, whether in a simple ratio or a more complex mathematical model, oversimplifies a complex issue. Fiscal soundness is a product of both physical and behavioural realities. Oversimplification introduces selection bias and affects the very conceptualisation of how fiscal soundness is analysed. This is because public deficits or surpluses are not simply a result of policy choices. The size of government deficits or surpluses also depends on the current and future state of the economy. Economic conditions, in turn, include not just the level and growth rate of national income but the health of all capital assets. As a result, fiscal sustainability is more usefully analysed by looking at all assets and debts that affect national income and economic capacity, as is further developed in the next section.

In the present chapter, fiscal sustainability is defined as the ability to continue current services and capital maintenance into the future, based on three factors: (a) the spending responsibilities of the political unit,[5] (b) its expected revenues, and (c) its ability to borrow at affordable rates of interest on the capital market. This definition is similar to one used by the GASB for state

and local governments in the US. It allows for changes in the structure or level of taxes, and it does not imply a balanced budget. The following section addresses how to establish indicators of fiscal sustainability that are based on the economy as a whole and why this approach is more likely to predict future fiscal sustainability.

The Basis of Fiscal Sustainability

The bedrock of fiscal sustainability is a sustainable and productive economy, and that depends on a nation's assets. Just as national income (GNI) reflects both public and private economic activity, national assets represent a combination of privately and publicly held stocks of many types of assets. Debt service and debt repayment will come from future income, but that income will be derived from the quantity and quality of capital stocks—human and natural assets as well as manufactured assets. Intangible assets such as the institutions and social capital of a society also contribute to the productivity of physical capital assets and the growth and stability of national income. The tax structure is one of these institutions, and its effectiveness determines whether a government can draw on national income to pay future expenses and to service debts that have been incurred.

However, the dominant economic model of the last century—neoclassical economic theory—has been focused on private (business) capital assets, with some attention to individual human capital or skills. In contrast, sustainable development theory considers public capital (infrastructure) and natural resource capital as equally important stocks of assets that underlie the capacity to produce output and income. In addition, political stability and the extent of corruption also influence how well resources are used in producing national income. These humanly developed institutions and social norms (sometimes called social capital) can be thought of as commonly held elements of human capital, somewhat analogous to the commonly held natural capital of earth's atmosphere. They are what economists call "pure public goods", because although the economy and society depend on them for their ongoing operation, they cannot be owned by any one person, and therefore cannot yield a private return (see Greenwood and Holt 2015, pp. 164–167; 2014, pp. 12–14).

Figure 14.1 illustrates how fiscal sustainability (the triangle at the top) depends on relationships between public revenue, public spending, and financial market confidence. Budgets do not always have to be balanced, particularly in sovereign nations, as long as there is the ability to borrow at reasonable

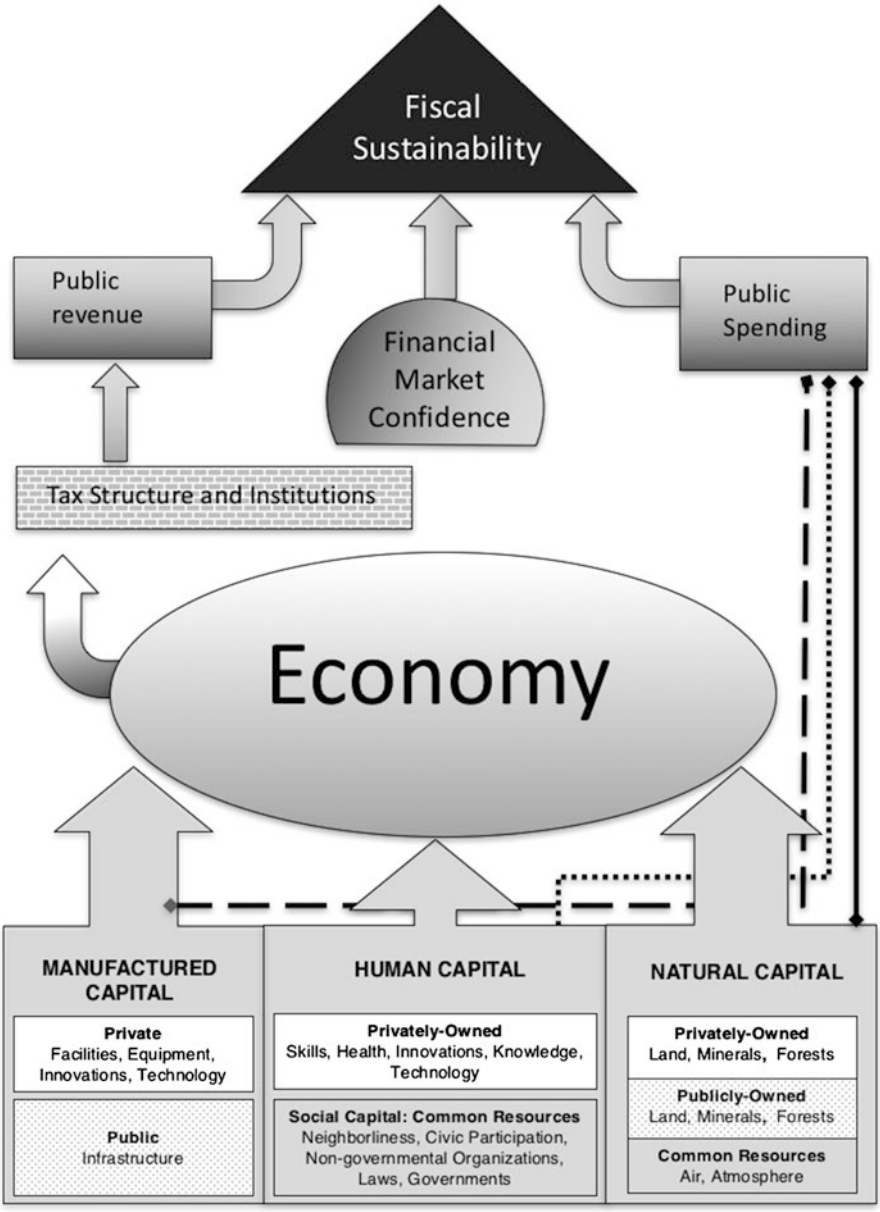

Fig. 14.1 Capital stocks underlie future revenues, spending, and fiscal sustainability. Source: Own design

rates in the financial markets. So far, this discussion reflects the traditional fiscal sustainability focus on revenue, spending, and market confidence. However, moving to the bottom of Fig. 14.1, the quantity and quality of capital stocks underlie both economic performance and public spending. And in the centre of Fig. 14.1, public revenue is determined in part by economic performance but is filtered through the tax structure, a public institution which is part of social capital. Two different societies with equal GNI per capita could have very different levels of public revenue due to widely different social preferences for private versus public goods or variation in administrative efficiency and compliance with their tax systems.

Another way to look at Fig. 14.1 is through five equations that summarise how social, economic, and environmental factors jointly determine fiscal sustainability in the public sector. Equation (14.1) reflects the top layer of Fig. 14.1: fiscal sustainability (FS) depends on public expenditure (X), public revenue (PR), and financial market confidence (FMC).

$$\mathrm{FS} = \mathrm{f}\left(\mathrm{PR}, \mathrm{X}, \mathrm{FMC}\right) \tag{14.1}$$

Next, public revenue (PR) depends on how the tax structure (TX) generates income for the public sector from a given level of national economic activity (E). Economic activity includes not only the level and growth rate of GNI, but also real per capita income and changes in how income is distributed (Persson and Tabellini 1994). Each of these can affect future economic activity, tax revenue, and public expenditure. In addition, the relationship between public revenue (PR) and the tax structure (TX) depends on the relative efficiency (or corruption) of public institutions, as shaped by the institutions (I) that are part of the shared social capital.

$$\mathrm{PR} = \mathrm{f}\left(\mathrm{E}, \mathrm{TX}, \mathrm{I}\right) \tag{14.2}$$

Turning next to the level, growth, and distribution of economic activity (E) in equation 14.3, we see that the economy depends on all the capital stocks—whether they are held privately, publicly (through government), or commonly. Privately held manufacturing capital (K) and human capital (H) are major contributors to economic activity. But production also requires publicly owned infrastructure (PI) and natural resources or assets (N). This "natural capital" may be publicly or privately owned. Some natural assets (forests, mineral reserves, etc.) are

the basis for economic activity, income, and tax revenues while others (wetlands) provide direct environmental services. Many natural assets, such as air and most of the ocean, are held in common and have no market-based economic value. However, if they provide necessary environmental services to society, depletion or degradation of them can lead to loss of human life, along with destruction of public and private economic assets.

Sustainable development theory also emphasises social capital, which is comprised of more than laws and governments, since it also includes many widely accepted norms and customs. These are all included in Fig. 14.1 under institutions (I). To function well, an economy needs the stability provided by customs and traditions, but there must be sufficient flexibility in these customs to allow social and economic progress. For example, long-established social traditions can require under-utilising female intellectual talents or block the adoption of new technologies. In fact, customs and traditions are clearly not all positive, as they include racial and ethnic discrimination which hampers economic development. Economies flourish with a balance between a stability of customs and traditions and an openness to change. Lastly, although not shown in Fig. 14.1, the level of knowledge and technology (T) that comes from human capital and the business cycle (B) affects the level and growth rate of economic activity.

$$E = f\left(K, H, PI, N, I, T, B\right) \tag{14.3}$$

Public expenditures (X) are part of the fiscal sustainability equation, and they also depend on more than policy decisions. They are affected by one-time (O) or temporary costs of macroeconomic stabilisation, disaster, or war, as well as by the quantity and quality of capital stocks. For example, the cost of public infrastructure construction and maintenance (PIc) will be influenced by environmental costs (Nc) from global warming or water contamination as well as environmental services from wetlands (Ns). There may also be higher public spending for disaster relief and/or mitigation. Future social costs (Sc) from disease, poor health, or insufficient education will also depend in part on the maintenance and improvement of many current capital stocks.

Lifecycle costs (Lc) include costs related to demographic transitions, such as the ageing populations of western Europe, Japan, and the US. These are expected to raise pension and medical costs. However, many developing countries have the opposite kind of lifecycle costs—young populations that require rising expenditure on public education. Since population dynamics change, tax revenues and spending burdens will also ebb and flow over time with demographic change. The last of the non-policy related elements driving future

expenditures is the business cycle (B), since more is spent on unemployment insurance and other social support programmes during economic downturns.

$$X = f(E, I, PIc, Sc, Nc, Ns, Lc, O, B) \tag{14.4}$$

To summarise the linkages shown in Fig. 14.1, revenue comes from the economy, but is filtered through the tax structure, an important social institution. The state of the economy (E) depends on the quantity and quality of capital stocks—privately manufactured, public infrastructure, human, social, and natural, and the level of technology. Capacity utilisation also matters, and it depends on many aspects of social capital as well as on appropriate stabilisation policies during business downturns. The indicators of E include not just the level of GNI but also its current and projected growth rates, as well as the way in which income and wealth are distributed among the population (Aghion et al. 1999; Persson and Tabellini 1994).

Public expenditure depends on many of the same capital stocks, technology, and institutions that affect economic performance. For example, better human capital in the form of improved health is likely to require less social spending per capita in the future. Conversely, failure to maintain built and natural infrastructure is likely to result in large unplanned public expenditures when natural disasters occur. High inequality of income or wealth raises social costs (Sc) if it is linked to poverty or deprivation, with fewer people able to independently meet their basic needs (Wilkinson and Pickett 2010). Along with the effects of an excessive concentration of income and wealth on economic growth, there are also effects on political participation (Solt 2008) and growth. In addition, the possibilities for corruption and disproportionate political influence grow as income and wealth become more concentrated (Glaeser et al. 2003; Jong-Sung and Khagram 2005).

However, fiscal sustainability depends not only on the assets outlined in Fig. 14.1 and the way in which they affect revenue and expenditure (Equations 14.1, 14.2, 14.3, and 14.4). It is also affected by the ability to borrow at affordable rates on the financial (credit) market. This ability is largely a function of investor confidence, which in equation (14.5) is also based on the strength of the economy, tax systems, and political and social institutions. In short, it all goes back to the same fundamental capital stocks—natural, human, privately manufactured, public infrastructure, and social (also called "institutions").

$$FMC = f(E, T, I) \tag{14.5}$$

A stunning example of how important economic capacity and political institutions are relative to public debt levels occurred in the US after the Great Recession. Based on rising debt/GNI in 2011, Standard and Poors downgraded the US bond ratings from AAA to AA (Goldfarb 2011). While a ratings downgrade generally raises interest rates for a private company, this one had no effect on the ability of the US government to continue issuing new bills and bonds at very low rates of interest. It appears that there was sufficient information available about the quality of the US capital stocks to maintain financial market confidence despite lower bond ratings. Subsequently, recovery in the US economy caused public revenue to grow and the national deficit to shrink. The debt/GNI ratio began to increase more slowly due to increases in the denominator (GNI) and smaller additions to the numerator (debt) as the gap between expenditure and revenue shrank. By 2013–2014, debt/GNI had stabilised in the US.[6] Revenue had been falling due to high unemployment and lower income and business profits, not because of an inadequate asset base of the capital stocks needed for production. Public spending on social safety net programmes had ratcheted up due to temporarily high unemployment. Thus, deficits and debt had both risen temporarily but could be reversed by countercyclical spending.

This example underscores the need for nations to adequately invest in all the capital stocks, including political institutions, to maintain confidence in financial markets as well as economic capacity. In the long run, tax structures that create a sustainable revenue stream are also necessary. Let us turn next to them and how they relate to fiscal sustainability and its measurement.

Tax Structures and Revenue Adequacy

Even with strong capital stocks and a healthy economy, revenue can be insufficient to cover expenditure if the tax structure does not effectively capture a portion of wage and productivity growth. The second foundation of fiscal sustainability is an efficiently and fairly administered tax structure that is flexible enough to accommodate economic change. Indicators of fiscal sustainability should reflect how well a tax system (e.g. a value-added tax vs. an income tax or a carbon tax) adapts to changing trends such as e-commerce, globalised production, or greater consumer spending on services relative to goods. Widespread co-operation with the tax system lowers enforcement costs per dollar of revenue and makes the tax system more efficient. This is more likely in a relatively open democracy where public sector corruption is low. For this

reason, indicators of corruption, trust in government, and the cost of tax collection are quite relevant to fiscal sustainability. This is another example of how non-monetised indicators can be quite useful in predicting trends in public revenue and financial market confidence.

However, tax revenue is affected by more than policy decisions and public co-operation. Since it depends on the level of economic activity, it varies in the short term with the business cycle. Therefore, long-term tax revenues will be higher when there are fewer recessions and when they are shorter and shallower. However, future tax revenues may also be influenced by demographic change in a variety of ways.[7] As affluent countries are experiencing ageing populations, substantial attention is now paid to how a rising share of retirees lowers the percentage of the population that pays labour-force-related taxes. When the analysis shifts out of a static model to a dynamic model with feedback effects, a smaller cohort of workers in a population is likely to raise real wages above what they are today, and contributions per worker are also likely to be higher.

To summarise, the levels and stability of future tax revenue are affected by economic activity, demographic change, tax policy, and public co-operation. Fiscal sustainability measures should reflect all of these. However, fiscal sustainability is also affected by expenditure, which the next section addresses.

Future Public Expenditure

Many public expenditure projections assume a "business as usual" model, with the possible addition of risks such as changing dependency ratios or unanticipated spending shocks from wars or disasters. However, current policy decisions regarding investments in human, natural, and physical capital stocks affect future spending in ways that are easier to predict. Even without policy changes in current programmes, future real spending per capita may be higher or lower, depending on the state of future capital stocks. This section focuses on the size and the quality of current capital stocks and the impacts of depletion or investment in them on future sustainability.

Let us begin with public infrastructure, the area most commonly thought of as "public capital". If many bridges, railways, roads, or airports are currently in poor condition, then it is much more likely that future expenditure on repairing or replacing them will be greater than it is currently. The same is true for the condition of wetlands that buffer major urban areas and croplands from storms and flooding (sometimes called "green infrastructure"). For both public infrastructure and natural capital, going beyond preservation to quality

improvement can lower per capita public expenditure in the future. For example, "smart" infrastructure that incorporates new technologies to prevent traffic accidents and reduce time delayed in traffic can reduce the public costs of dealing with accidents and increase private-sector productivity. Similarly, restored wetlands can lower flood damage from future storms.

When it comes to future public spending on medical care, public safety, and education, these also depend on more than eligibility and programme criteria. Economists have focused on human capital as a vital contributor to productivity growth and cautioned that sufficient investment in health and education is needed to sustain productivity and economic growth. However, the quality of human capital also determines the need for many types of social spending.[8] For example, high-quality pre-school for low-income children (Bartik 2011; Heckman and Carneiro 2003) has been credited not only with increasing future earnings but also with decreasing costs in the social service and criminal justice systems in the future (Waldfogel and Washbrook 2011).[9]

Clearly, the quantity and quality of capital stocks affect more than economic capacity. Future expenditure per capita may be higher or lower depending on whether capital is depleted, maintained, or increased through investment. A thorough approach to fiscal sustainability requires moving beyond purely fiscal measures in order to consider fiscal relationships to economic, environmental, and social sustainability. Local indicators from around the world provide examples of how to use a combination of monetary and non-monetary indicators to track sustainable development. In the next section, some of these indicators are examined, along with their relationship to fiscal sustainability.

Local Indicators of Sustainable Development

The economic future depends on a variety of capital stocks as outlined in Fig. 14.1. Some are privately owned, some are publicly owned, and some are commonly held. Local indicator projects go beyond business investment in the national income accounts to include measures that indicate whether human capital, public infrastructure, and the environment are being sustained for the future. They look at social capital through measures of voter turnout, trust in government, crime, and ethnic or racial division to track changes in social cohesion. Many include some fiscal measures, such as tax capacity.

Introduction to Local Indicators

Local sustainability and quality-of-life indicators originated in English-speaking countries such as Australia, Canada, Great Britain, and the US (Greenwood 2004; Greenwood and Holt 2015), but they are being developed by many cities around the world.[10] Measures have been identified that track sustainability or quality of life in ways that standard economic measures do not, by addressing aspects of economic, environmental, and social capital. Many projects began in non-governmental organisations or universities but have since been absorbed by formal governments. A variety of measures are used in order to better reflect quality of life and sustainable development than GNI per capita or its growth rate does. While some indicators relate to current activity, others reflect a wide range of debts and assets. None alone can give an accurate reading on sustainability, but taken as a whole a set of indicators can show improvement, stability, or decline.

These local indicators—although chosen to show progress towards sustainable development—provide a starting point for understanding the fiscal sustainability of governments. The best projects look at the economy, environment, and society holistically and emphasise linkages and trade-offs between various indicators—an approach that fiscal sustainability must take to go beyond debt-to-income ratios. A great deal of the data used in local indicator projects come from existing administrative or survey data, often collected at the national level. The data is rarely original—it is how the elements of data are used that is innovative.

Let us first look at indicators chosen to reflect the current state of the economy. Housing affordability is an indirect indicator of income distribution, because it reflects whether average pay is keeping pace with increases in housing costs. Poverty rates are another readily available indicator. Both poverty rates and income distribution are current measures but they are likely to affect future economic performance, and therefore will also affect public revenue and public spending. This is particularly true for child poverty rates. Another indicator of economic sustainability used by local governments is the reliance of job creation on a few sectors. The more jobs are concentrated in a few sectors, the greater the economic vulnerability to downturns in economic activity and to long-term change. This indicator could be a very appropriate signal of fiscal sustainability for small or developing countries.

Examples of Local Indicators

Good indicators accurately reflect what is happening and communicate it effectively to the public and to leaders. No single indicator can measure an entire capital stock, but some are like the proverbial "canary in the coal mine", the miners' early warning system of unsafe air. Specific choices of indicators are often based on data availability, local problems or issues, and what is understandable to concerned citizens (Greenwood 2004). The best indicators are forward-looking, helping to predict future trends. Table 14.1 shows examples of local indicators of sustainable development from across the world that

Table 14.1 General sustainability indicators that underlie fiscal sustainability

Concept	Indicator	Example of use
Economic capacity	Under-employment	Jacksonville, FL, US
	Adults with BA/BSc degree or higher	Jacksonville, FL, US
	Poverty rate	Austin, TX, US
Economic sustainability	% of green jobs in the economy	Portland, OR, US
	% of total new businesses surviving 3+ years	Austin, TX, US
		Austin, TX, US
	% of total jobs in the public sector	Seattle, WA, US
	Child poverty rate	
Tax capacity	Housing Affordability – Repayments as % of Income	Dublin, Ireland
		Jacksonville, FL, US
	% of households spending over 30% of income on housing	Austin, TX, US
	Taxes as % of average income	
Natural capital/ Green infrastructure	Level of reforestation	Bogota, Colombia
	Percentage area of tree canopy	Portland, OR, US
	Ecological status: water bodies	Dublin, Ireland
	Percentage of land surface impervious to water	Austin, TX, US
		Seattle, WA, US
Public infrastructure	% of roads in good condition	Bogotá, Colombia
	Improved traffic flow	Tel Aviv, Israel
	Energy efficiency of public buildings	Buenos Aires, Argentina
		Dublin, Ireland
		Tel Aviv, Israel
Human capital	Public school dropout rate	Ilheus, Brazil
	High school graduation rate	Jacksonville, FL, US
	% adults at healthy weight	Portland, OR, US
	% 8th graders at healthy weight	Seattle, WA, US
	Babies with low birth weight	Ilheus, Brazil
Social capital	% reporting trust in city leaders or government	Jacksonville, FL, US
		Austin, TX, US
	Crime rate	Seattle, WA, US
	Racial disparities in juvenile courts	Austin, TX, US

Sources: Greenwood (2004) and Sustainable Cities International (2012)

also relate to fiscal sustainability. The relationship might be through the strength of the economy, the stability of the underlying political-social structure, or the future expenses and revenues of governments regarding infrastructure, human populations, or the natural environment.

Sources of Information for Constructing Capital Stock Indicators

Local indicator groups use both administrative data and surveys, many of which are collected nationally. For example, in the US local poverty rates are computed by the U.S. Census Bureau (n.d.) from family income information in annual American Community Surveys and other sources. Schools in the US report high school graduation rates and the percentage of students who drop out of school to the U.S. Department of Education (n.d.). Since youth development is so important to future economic capacity (and to future social spending), these are some of the most important numbers for fiscal sustainability. The percentage of babies born at low birthweight is collected by health agencies in most countries and is another very important indicator of problems in human capital development (United Nations Children's Fund and World Health Organisation 2004).

For natural capital, air and water quality measures are compiled by the US Environmental Protection Agency (EPA n.d.) and the U. S. Geological Survey (2016), but many other measures come from information collected by universities or non-governmental organisations. For example, measures of tree canopy coverage are compiled in the US by non-governmental organisations such as the University of California, Davis (n.d.). Urban development has led to rapid increases in the percentage of land area impenetrable by water—often called impermeable surfaces—which directly increases risks of flooding and can damage water quality and fish populations. Flooding and water quality will affect future public expenditure, while fish populations will affect economic activity and public revenue in some areas. Many local indicator projects measure impermeable surfaces because they are concerned with future expenditures. They previously relied on ground surveys and aerial photographs but are now supplementing this information with less expensive satellite imagery (University of Minnesota n.d.)

The quality of public infrastructure, such as roads, bridges, and dams, is graded annually in the US by a private non-profit association, the American Society of Civil Engineers (2013). Infrastructure maintenance backlogs are also available in the public reports of individual states and municipalities due

to new rules treating these backlogs as public debts (see the section below). One element of public infrastructure today will directly affect future operating costs—the energy efficiency of existing and newly constructed public buildings. Measuring this has been addressed by the United Nations Environment Programme (2010) and in the US by State and Local Energy Efficiency Action (2014), among others.

Social capital is probably one of the most difficult areas to measure, and its indicators come primarily from public surveys. These include questions on trust in government, elected officials or law enforcement, as well as on perceived public safety. However, trends in violent crime rates (US Federal Bureau of Investigation 2014) and disparities in the treatment of different ethnic or racial groups in the justice system (American Civil Liberties Union 2016) can also shed light on changes in social cohesion.

Tax capacity, an important aspect of fiscal sustainability, depends on more than income growth and the share of income going to taxes, although both are commonly measured in these projects. Comparing housing costs to family income gives an idea of how much discretionary income people have after meeting their basic needs. Indicator projects combine average rents and housing prices with pay or income levels to assess housing affordability. The same information can also be used as an indicator of discretionary income, and hence tax capacity. While there is no public data on rental costs in the US, the National Multifamily Housing Council (2016) makes its data publicly available. From the examples above, a great deal of information is already collected by different levels of government that can be used to assess changes in capital stocks. By combining it with reports from non-governmental organisations, universities, and information gathered by private trade associations, new light can be shed on sustainable development and the fiscal sustainability that goes with it.

Measures of Fiscal Health for Sub-national Governments

While local indicators of sustainability or quality of life often contain some fiscal measures, internal reports on government performance, such as the accounting statements released to the public and investors, are a richer source for public fiscal measures. The fiscal standards promulgated by the GASB for the private and the public sector in the US focus on non-financial debts and assets as well as financial ones. For example, the public sector standards have been modified to require disclosure of the infrastructure maintenance backlog on local government balance sheets, since

these are postponements of a financial obligation (GASB 2008, 2011). Governments are also now asked to annually report the total future cost of taxes waived to attract or retain businesses as long as the tax abatements continue, and this relates to tax capacity (see GASB 2014, pp. 2–10).[11]

Capital stock measures for national governments could be similar in some cases to those chosen by local governments, but would be likely to be quite different in others. Accounting methods for sovereign governments vary considerably from those for sub-nationals. However, cities and other local governments provide innovative examples of capital stock indicators that could inspire better indicators and accounting methods to avoid relying exclusively on GNI and debt/GNI as indicators of fiscal health and sustainability.

Putting It All Together

Over the long term, fiscal sustainability depends on a productive economy more than on any financial measure. It requires (a) an economy based on high-quality manufactured, natural and human capital stocks, (b) a "climate of innovation" that is open to technological and institutional change while preserving stability and social cohesion, and (c) a lack of corruption. The risk of large future deficits depends on relative changes in public revenue and public expenditure. Revenues are a function not only of potential national income, but also of full employment policies and a fair and efficient tax system. Levels of public spending depend not only on policy choices, but on changing needs for infrastructure and changing human needs. The needs of the future depend in a significant way on the investments made in the present.

When countries need to borrow to deal with temporary catastrophes (war, natural disasters, etc.), stabilisation over the business cycle or to fund public investments, the cost of borrowing (or the interest rate) depends on general financial market confidence in the country's ability to repay its debts in the future. That confidence is based on the very same factors discussed above: capital stocks, stability, and openness to technological change which positively affect economic growth, filtered through an economic and political system that allows for repayment of sovereign debts through taxation.

While under particular circumstances the ratio of debt to income may point to a fiscal problem, the *ad hoc* nature of this measure limits its general usefulness. For example, debt to income did not prove useful in anticipating the fiscal crises in EU nations (such as Spain) that did not have high government debt in 2008–2009.[12] Even broader debt measures that included private debt (the

problem in Spain during the financial crisis) would still be purely financial measures, omitting the underlying importance of capital stocks and social institutions to fiscal sustainability.

In conclusion, five key principles about fiscal sustainability and its measurement have been identified in this chapter.

1. A single measure will inevitably fail to capture the complexity of economic systems and fiscal realities and raises real problems of selection bias. A more comprehensive set of indicators, largely based on the health of capital stocks, is needed to supplement the commonly used indicator of the debt/GNI ratio.
2. Relying on an incorrect or incomplete measure of fiscal sustainability provides the wrong conceptual framework for policy and can interfere with adequately maintaining the level of economic activity, as well as with investment in "green" infrastructure, human capital, and public safety.
3. As a result, pressure to lower the public-debt-to-income ratio can actually *increase* the risks to long-term fiscal sustainability. For example, if a government maintains its high bond rating by neglecting to spend funds on its natural and built infrastructure, and subsequently experiences devastation from a natural disaster (such as a flood) or an epidemic (such as Zika or Ebola), this could endanger fiscal sustainability by requiring future spending that could easily be far in excess of the cost of preventive measures.
4. Public debt levels could be more constructively evaluated if national governments maintained separate capital, operating, and stabilisation accounts. When these are combined, the entire budget tends to be viewed as an operating budget, without recognition of stabilisation or capital investment spending. Yet both stabilisation and capital investment will produce higher income in the future, raising the denominator of the debt-to-income ratio along with the numerator.
5. Fiscal sustainability is not an end in itself, but a *means* of maintaining or improving the quality of life of citizens in the future. This is another reason for considering it within the context of economic, environmental, and social sustainability and using a variety of indicators, including many taken from the sustainability literature.

This chapter has presented arguments that (a) fiscal sustainability is best understood within the framework of sustainable development, and (b) capital stock indicators should be part of fiscal sustainability assessments at the national level. Despite its lack of predictive power, debt/GNI retains a hold on the public imagination and the minds of policy makers. However, fiscal

sustainability assessments could be more accurate if additional measures that are grounded in sustainable development theory were added. The indicators of sustainability presented earlier in Table 14.1 provide examples of how the health of various capital stocks can contribute to a richer understanding of economic, environmental, and social sustainability. Many of these would also meet the important requirement of being understandable to non-specialists (as identified by Bhuta et al. 2014).

Notes

1. There is substantial controversy over whether using shadow prices to monetise capital stocks like health and the environment, as in the IWI, adds or subtracts from precision. The study by Bhuta et al. (2014) concluded that the results do not "provide a clear image of resource scarcity or depletion" (p. 11).
2. The work of Rogoff and Reinhart (2010) is often cited to support concern over high ratios of debt/GNI and potential damage to future economic growth. However, their 2010 paper showed very weak relationships in established and affluent economies between debt/GNI ratios of less than 90% and real economic growth. Subsequent critiques of Reinhart-Rogoff based on recalculations of the same data (Herndon et al. 2014) showed that the 90% ratio clearly did not apply to the 20 most affluent nations.
3. The IMF (2000) report goes on to discuss the difficulty of identifying critical ranges for debt indicators "across heterogeneous countries without additional information" such as the average interest rate, the country's tax base, the pace of output and export growth, the composition of debt, exchange rate vulnerability, and indicators of general corporate profitability (point 63).
4. GASB (2011) adds the proviso that costs should not be shifted among generations. This presumably refers to shifting current operating costs into the future through debt financing, since state and local governments in the US regularly use debt to finance infrastructure that yields benefits to future generations. Note that this constraint is potentially much broader than its current application to purely financial matters. If the present-day generation uses fuels that emit carbon and impose costs on future generations, this should be included under the GASB definition.
5. The unit may be a nation-state or one of a whole variety of sub-national governments, such as cities, counties, provinces, departments, cantonments, etc.
6. The best source for debt levels in the US over time is Treasury Direct (2017).
7. Higher wages are also likely to attract new workers into the labour force, including immigrants from abroad, which will somewhat offset the declining worker-to-population ratio.

8. Public health and preventive medicine represent a particularly important area where current investment can reduce future per capita spending. Since medical expenditure for the elderly absorbs a large part of public budgets in most affluent countries, a multifaceted approach to "health" that includes lifestyle modification could improve health at the same time reduces future spending for the elderly population than trend forecasts currently predict. This would free up future public revenue for other types of spending, including debt service or repayment. Better public health and preventive medicine could also positively impact child development and workforce health, contributing to higher labour productivity and income, along with lower medical costs.
9. Waldfogel and Washbrook (2011) also credit widely available and high-quality pre-school programmes in Great Britain with Great Britain's higher economic and social mobility relative to the US (despite high income inequality and child poverty in both countries).
10. The Global City Indicators Program (GCIP) funded by the World Bank has nine pilot projects, including Sao Paulo, Bogota, Cali, Belo Horizonte, Porto Alegre, and Toronto (The Global City Indicators n.d.)
11. GASB (2014) is consistent with the literature on location incentives, which shows that rates of return on these are consistently low to non-existent. This suggests that fiscal sustainability would be far better served by spending on infrastructure.
12. Spain had a relatively low public debt/GNI, although its high private debt (incurred largely through real estate speculation) led to economic problems once there was a downturn in the real estate market. As with Greece, Spain's inability to print its own currency contributed to the capital market issues. For more, see Krugman (2010).

References

Aghion, P., Caroli, E., & Garcia-Penalosa, C. (1999). Inequality and Economic Growth: The Perspective of the New Growth Theories. *Journal of Economic literature, 37*(4), 1615–1660. doi:10.1257/jel.37.4.1615.

American Civil Liberties Union (ACLU). (2016). *Racial Disparities in Florida Safety Belt Law Enforcement.* ACLU. Retrieved from https://www.aclu.org/report/racial-disparities-florida-safety-belt-law-enforcement

American Society of Civil Engineers (ASCE). (2013). *Report Card for America's Infrastructure*. ASCE. Retrieved from http://www.infrastructurereportcard.org/

Bartik, T. (2011). *Investing in Kids: Early Childhood Programs and Local Economic Development*. Kalamazoo: Upjohn Press.

Bhuta, N., Malito, D. V., & Umbach, G. (2014). *Representing, Reducing or Removing Complexity: Indicators of Sustainability and Fiscal Sustainability* (Robert Schuman Centre for Advanced Studies (RSCAS) Working Paper, 2014/78). Florence: European University Institute. Retrieved from http://www.eui.eu/RSCAS/Publications/

Bruntland, G. H. (1987). *Our Common Future: Report of the World Commission on Environment and Development*. Oxford: Oxford University Press.

Chalk, N., & Hemming, R. (2000). *Assessing Fiscal Sustainability in Theory and Practice* (International Monetary Fund Working Paper Series WP/00/81). Washington, DC: International Monetary Fund (IMF).

Congressional Budget Office. (2010). *Fiscal Stress Faced by Local Governments*. Washington, DC: Congressional Budget Office.

European Commission, Directorate General for Economic and Financial Affairs. (2012). *Fiscal Sustainability Report 2012.* Brussels: European Commission. doi: 10.2765/19669.

GASB. (2008). *Exposure Draft of an Improved Conceptual Framework for Financial Reporting*. Norwalk: GASB. Retrieved from http://www.ifrs.org/Current-Projects/IASB-Projects/Conceptual-Framework/EDMay08/Documents/conceptual_framework_exposure_draft.pdf

GASB. (2011). *Preliminary Views on Major Issues Related to Economic Condition Reporting: Financial Projections*. Norwalk: GASB. Retrieved from http://www.gasb.org/cs/BlobServer?blobkey=id&blobwhere=1175823466973&blobheader=application%2Fpdf&blobcol=urldata&blobtable=MungoBlobs

GASB. (2014). *Exposure Draft: Tax Abatement Disclosures.* Norwalk: GASB. Retrieved from http://www.gasb.org/jsp/GASB/Document_C/GASBDocumentPage?cid=11761644979

Glaeser, E., Scheinkman, J., & Shleifer, A. (2003). The Injustice of Inequality. *Journal of Monetary Economics, 50*(1), 199–222. doi:10.1016/S0304-3932(02)00204-0.

Goldfarb, Z. A. (2011, August 6). S&P Downgrades U.S. Credit Rating for First Time. *The Washington Post*. Retrieved from https://www.washingtonpost.com/business/economy/sandp-considering-first-downgrade-of-us-credit-rating/2011/08/05/gIQAqKeIxI_story.html

Goodwin, N. R. (2003). *Five Kinds of Capital: Useful Concepts for Sustainable Development* (Global Development and Environment Institute Working Paper No. 03–07). Medford: Tufts University.

Government Accounting Standards Board (GASB). (2007). *The Users' Perspective*. Norwalk: GASB. Retrieved from http://www.gasb.org/jsp/GASB/GASBContent_C/

Greenwood, D. T. (2004). Measuring Quality of Life with Local Indicators. In E. N. Wolff (Ed.), *What Has Happened to the Quality of Life in the Advanced Industrialised Nations?* (pp. 334–374). Cheltenham: Edward Elgar Press.

Greenwood, D. T. (2010, November 4). *The Debts and Assets of Future Generations: Looking Beyond the Financial.* Presented at the Intergenerational Well-being Seminar, Kraemer Family Library. Colorado Springs: University of Colorado.

Greenwood, D. T., & Holt, R. P. F. (2014). Improving the Standard of Living: Income, Quality of Life and Sustainability. In R. P. F. Holt & D. T. Greenwood (Eds.), *A Brighter Future: Improving the Standard of Living Now and for the Next Generation* (pp. 3–22). Armonk: M.E. Sharpe, Inc.

Greenwood, D. T., & Holt, R. P. F. (2015 [2010]). *Local Economic Development in the 21st Century: Quality of Life and Sustainability*. New York: Routledge.

Heckman, J. J., & Carneiro, P. M. (2003). *Human Capital Policy* (National Bureau of Economic Research, NBER, Working Paper No. 9495). Washington, DC: NBER.

Hendrick, R. (2004). Assessing and Measuring the Fiscal Heath of Local Governments. *Urban Affairs Review, 40*(1), 78–114.

Herndon, T., Ash, M., & Pollin, R. (2014). Does High Public Debt Consistently Stifle Economic Growth? A Critique of Reinhart and Rogoff. *Cambridge Journal of Economics, 38*(2), 257–279. doi:10.1093/cje/bet075.

International Monetary Fund (IMF). (2000). *Debt- and Reserve-Related Indicators of External Vulnerability*. Washington, DC: IMF. Retrieved from http://www.imf.org/external/np/pdr/debtres/

International Public Sector Accounting Standards Board (IPSASB). (2011). *Recommended Practice Guideline, Reporting on the Long-Term Sustainability of a Public Sector Entity's Finances*. New York: International Federation of Accountants. Retrieved from https://www.ifac.org/publications-resources/recommended-practice-guideline-reporting-long-term-sustainability-public-sect

Jong-Sung, Y., & Khagram, S. (2005). A Comparative Study of Inequality and Corruption. *American Sociological Review, 70*(1), 136–157. doi:10.1177/000312240507000107.

Krugman, P. (2010, February 5). *The Spanish Tragedy*. Retrieved from https://krugman.blogs.nytimes.com/2010/02/05/the-spanish-tragedy/?_r=1

National Multifamily Housing Council (NMHC). (2016). *Data Sources: Apartment Rents (US)*. Washington, DC: NMHC. Retrieved from http://www.nmhc.org/Content.aspx?id=4418

Persson, T., & Tabellini, G. (1994). Is Inequality Harmful for Growth? *The American Economic Review, 84*(3), 600–621. Retrieved from http://www.jstor.org/stable/2118070

Rogoff, K., & Reinhart, C. (2010). Growth in a Time of Debt. *American Economic Review, 100*(2), 573–578. doi:10.2469/dig.v40.n3.19.

Solt, F. (2008). Economic Inequality and Democratic Political Engagement. *American Journal of Political Science, 52*(1), 48–60. Retrieved from http://www.jstor.org/stable/25193796

State and Local Energy Efficiency Action Network. (2014). *Greater Energy Savings Through Building Energy Performance Policy: Four Leading Policy and Program Options*. State and Local Enercy Efficiency Action Network. Retrieved from https://www4.eere.energy.gov/seeaction/publication/greater-energy-savings-through-building-energy-performance-policy-four-leading-policy

Sustainable Cities International. (2012). *Indicators for Sustainability: How Cities Are Monitoring and Evaluating Their Success*. Vancouver: Sustainable Cities International. Retrieved from http://www.mayorsinnovation.org/images/uploads/pdf/2__International_Case_Studies.pdf

The Global City Indicators. (n.d.). Website. Retrieved from http://www.cityindicators.org/

Treasury Direct. (2017). *The Debt to the Penny and Who Holds It*. Retrieved from http://www.treasurydirect.gov/NP/debt/current.

United Nations Children's Fund (UNICEF) and World Health Organisation (WHO). (2004). *Low Birthweight: Country, Regional and Global Estimates*. New York: UNICEF. Retrieved from http://data.unicef.org/resources/low-birthweight-country-regional-and-global-estimates.html

United Nations Environment Programme (UNEP). (2010). *Climate Friendly Buildings and Offices – A Practical Guide.* Nairobi: UNEP. Retrieved from http://www.unep.fr/scp/publications/details.asp?id=DTI/1278/PA

United States Census Bureau. (n.d.). *Poverty*. Washington, DC: United States Census Bureau. Retrieved from http://www.census.gov/topics/income-poverty/poverty.html

United States Department of Education. (n.d.). *Student Outcomes / Graduation Rates*. Washington, DC: US Department of Education. Retrieved from http://www2.ed.gov/rschstat/catalog/student-outcomes.html

United States Environmental Protection Agency. (n.d.). Washington, DC. Retrieved from https://www3.epa.gov/

United States Federal Bureau of Investigation. (2014). *Crime in the United States*. Washington, DC: United States Federal Bureau of Investigation. Retrieved from https://ucr.fbi.gov/crime-in-the-u.s/2014/crime-in-the-US-2014

United States Geological Survey (USGS). (2016). *USGS Water Quality Information*. Reston: USGS. Retrieved from http://water.usgs.gov/owq/data.html

University of California, Davis. (n.d.). *Tree Canopy Cover.* Retrieved from https://indicators.ucdavis.edu/forest/indicators/tree-canopy-cover

University of Minnesota, Remote Sensing and Geospatial Analysis Laboratory. (n.d.). *Impervious Surface Classification*. Retrieved from http://land.umn.edu/methods/imperv_class.html

Waldfogel, J., & Washbrook, E. (2011). Income-Related Gaps in School Readiness in the United States and the United Kingdom. In T. Smeeding, R. Erikson, & M. Jantti (Eds.), *Persistence, Privilege, and Parenting: The Comparative Study of Intergenerational Mobility* (pp. 175–207). New York: Russell Sage Foundation.

Wilkinson, R., & Pickett, K. (2010). *The Spirit Level: Why Equality Is Better for Everyone*. London: Penguin Books.

Part III

Applying Measures: The Use and the Impact of Indicators as Instruments of (Global) Governance

15

Corruption Indicators in Local Political Landscapes: Reflections from Albania

Smoki Musaraj

Introduction

Following the unparalleled success of Transparency International's (TI's) Corruption Perception Index (CPI) in the late 1990s, today a myriad of corruption indicators compete for donor funding, public recognition, and endorsement by policy-makers. A global assemblage of experts, international agencies, independent NGOs, consultancies, local and international politicians, researchers, and policy-makers is engaged in the production and circulation of these indicators. Rankings and assessments by these indicators play a significant role in the countries of the global South. In many countries, for instance, the corruption "grade" assigned by a corruption indicator often determines the level of international aid they receive and the possibility of political membership in regional blocs (see, for instance, Dutta 2015). Recent work by an interdisciplinary community of scholars has brought attention to the politics and economics of the production and deployment of indicators on a wide range of issues from human rights and the rule of law to health care and climate change (Davis et al. 2012a; Merry et al. 2015; Rottenburg et al. 2015; Merry 2016). Working alongside this literature, the current chapter looks into the production and circulation of a corruption indicator, mapping the heterogeneous network of actors that legitimise it and use it in both intended and unintended ways.

S. Musaraj (✉)
Ohio University, Athens, OH, USA

D.V. Malito et al. (eds.), *The Palgrave Handbook of Indicators in Global Governance*,
https://doi.org/10.1007/978-3-319-62707-6_15

The focus of the chapter is Albania, a country that over the past 25 years has depended on international development aid and has eagerly (but to this day unsuccessfully) tried to join the European Union (EU). Albania is also a country with consistently weak corruption "grades", a factor that is repeatedly underscored by international institutions such as the International Monetary Fund (IMF) and the World Bank (WB), as well as global political leaders such as the United States (US) and the European Union (EU). This chapter traces the production and circulation of a corruption perception survey that informs these national and international concerns about corruption in Albania. The indicator in question is "Corruption in Albania: Perceptions and Experience", produced annually between 2005 and 2010 by the local market research institute, the Institute for Development Research and Alternatives (IDRA), and funded by the United States Agency for International Development (USAID).[1] By looking at the production and circulation of several editions of the indicator (2008, 2009, 2016), the chapter addresses the following questions: How are indicators used by various actors? How does the motivation for the use of corruption indicators vary among these actors? How does the political economy of indicators (funding and institutional involvement) matter in gauging their impact on local political debate and decision-making?

The chapter begins with an outline of a network of actors—governmental bodies, international organisations, local research centres, and global consultants—involved in the production of the USAID/IDRA corruption perception survey. It situates the production of this indicator at the intersection of global and local histories of anti-corruption movements. Second, the chapter looks into a few key cases of high-level corruption brought to the courts between 2008–2010 and a more recent (2016) debate around judicial reform. As the chapter retraces the debates and decision-making processes around these court cases and reforms, it highlights the role of the USAID/IDRA indicator as a source of evidence and expertise and as a political device to direct policy. The chapter concludes with a discussion of the present debate on judicial reform and the intended and unintended role that corruption indicators continue to play in local political landscapes.

Global and Local Ecologies of a Corruption Indicator

"Corruption in Albania" was produced annually between 2005 and 2010 by the local market research centre IDRA with USAID Rule of Law Program funding. Locally, the survey was known as the USAID corruption indicator,

although IDRA conducted the bulk of the production process. The survey was discontinued in 2010 and resumed again in 2015–2016.

Although not part of a broader cross-country index (such as, for instance, the rival CPI by TI), like other corruption indicators the USAID/IDRA survey draws on what might be described as a global type of expertise on corruption and governance. This expertise entails (a) engaging in a global network of experts and organisations that measure and govern corruption; and (b) using a specific methodology (measuring perceptions using statistical and survey methods). The USAID/IDRA survey reflects both of these elements. Yet, it also illustrates how local actors assert their agency in the process of production of corruption and governance indicators.

Intersecting Histories and Networks of Anti-corruption

The USAID/IDRA corruption survey is situated within the broader context of an anti-corruption assemblage emerging in the mid-1990s, a network of "global integrity warriors" (Sampson 2005), an "industry" (Sampson 2010) consisting of experts, programmes, institutions, and large amounts of funding "to combat corruption" (Sampson 2010, p. 267). Debora Valentina Malito (2014) provides a comprehensive overview of the history and institutions that have come to populate this space. As she describes in detail, the target measured by these indicators extends well beyond corruption, focusing more broadly on governance and state capacity. Malito distinguishes between three types of governance indicators: survey-based perception indices, global governance indices, and state capacity indices. The USAID/IDRA survey falls into the category of survey-based indicators, which tackle corruption more specifically. Further, Malito lists the various sponsors of these indicators, which include a wide range of institutions from independent global research institutes (such as TI, Gallup International, and the Brooking Institute) to international organisations (the World Bank, the European Bank for Reconstruction and Development), to national and regional political agencies (the European Commission, the Canadian International Development Agency), to private and public universities (Columbia University, Michigan State University, George Mason University) (Malito 2014, p. 3). The funding and framing of the USAID/IDRA survey mirrors these global trends. At the same time, the production of this survey is shaped by the intersection of the global anti-corruption movement and local political networks and debates about corruption and state capture.

The USAID/IDRA survey emerged at the peak of a local anti-corruption movement that began in the late 1990s (Kajsiu 2013), after the political crisis and general anarchy following the collapse of a dozen pyramid schemes in the infamous year of 1997 (Musaraj 2011; Abrahams 2015). The Socialist Party (Partia Socialiste or PS) came into power in 1998 after two terms of the Democratic Party (Partia Demokratike or PD). By the end of the millennium, the PS-led government came under criticism for its high levels of corruption. Corruption became a target of civil society protests, and of the 2001 and 2005 PD electoral campaigns. The anti-corruption campaign was promoted and financed by various international institutions, including international organisations such as the World Bank, country-specific development agencies (e.g. the USAID), and regional bodies such as the EU and the Council of Europe (see also Kajsiu 2013, 2016).

The USAID/IDRA survey was one of the many anti-corruption initiatives sponsored at this time under the umbrella of the USAID Rule of Law Program. As Kajsiu (2013) documents, the USAID hired a Washington-based consultancy firm called Management System International (MSI) to help mobilise these anti-corruption initiatives. With the help of MSI, the Albanian Coalition Against Corruption (ACAC) was founded in 2001. This network then supported a number of local initiatives that aimed to raise awareness of and to measure corruption. Among other initiatives, the USAID/IDRA survey was funded to provide an overview of perceptions of corruption in the country and to serve as a sort of self-auditing tool for the work of the ACAC (see Seligson and Baviskar 2006). While the ACAC dissolved in the years to come, the USAID/IDRA survey continued to be produced. By 2008, the beginning of the ethnographic research that informs this chapter, the USAID/IDRA survey had become the most important indicator on corruption in the local public sphere, and was often considered even more reliable and truthful than the rival TI's CPI.[2]

As I describe in more detail below, the local anti-corruption efforts at this time centered on the work of then public prosecutor, Ina Rama, who (unsuccessfully) sought to prosecute three cases of high-level corruption. Rama's efforts were explicitely supported by then American Ambassador John Withers. While not effective at bringing to justice responsible officials, the anti-corruption efforts of 2008–2011 were effective in shifting the focus of the discourse of corruption (Gupta 1995) from petty corruption to state capture. Further, the target of this movement shifted from public awareness campaigns to prosecution of specific public officials and the denounciation of specific public procurement contracts.

The USAID/IDRA survey was discontinued in 2010 because of an end of the USAID Rule of Law Program. IDRA ran one issue of the survey with its own funds. Recently, IDRA resumed the survey with funding support from the US Embassy in Tirana (IDRA 2016a). The 2015–2016 survey contained a similar format and list of questions to the earlier surveys. This continuity allowed the survey to draw comparisons over time—a key feature of governance indicators. In addition to the corruption perceptions' survey, IDRA has conducted a number of other surveys that focus on related topics such as perceptions of impunity (IDRA and FSHSH 2015, funded by the Open Society Foundation's local chapter) and judicial reform (IDRA 2016b, funded by the Open Society Foundation and the US Embassy).

The restart of the IDRA survey in 2015–2016 came at a time of a renewed public debate around corruption, this time centered on judicial reform. PS has returned to power since 2013. In this third stage of anti-corruption movements in Albania, the dicourse of corruption continues to target high-level corruption; however, rather than specific officials, here the target is the whole justice sytem. As was the case in the two previous moments, this third moment of local anti-corruption campaigning also intersects with international anti-corruption discourse. In this case, the discourse of corruption and anti-corruption is more specifically framed by the European Union and its conditionalities around accession. Meanwhile, the most vocal political actors pushing for reform include government officials (such as Prime Minister Edi Rama, leader of PS) as well as international representatives (such as US Ambassador Donald Lu and EU Ambassador Romana Vlahutin).

The case for reform is made primarily on account of the country's poor record (or, rather, the lack of) prosecutions of high-level corruption. This time, the local anti-corruption campaign is waged by incumbents and framed within the broader context of application for European Union membership, a political aspiration shared by all political parties and local publics. Local oponents and critics, however, fear that the reform would be used primarily as a revenge tactic against the opposition party or as simply an obfuscation of the system rather than an effective and unbiased institutional reform. For European Union authorities, in turn, concerns about justice institutions in Albania are informed by more recent experiences and frustrations with corruption in Romania and Bulgaria (see also Serban 2015). Thus, while local actors use the European Union accession as the carrot for judicial reform, international actors use deferral of accession as a stick to punish Albania for its poor corruption "grade". In this carrot and stick game between international representatives and local politicians, corruption indicators constitute important data that measure corruption.

What we learn from these intersecting local and global histories of anti-corruption movements and policymaking is that a heterogeneous "ecology" of institutions (Halliday 2012) is mobilised in the production of governance indicators. This ecosystem cuts across local and global networks, and private, public, and public-private forms of funding. The process of designing and producing the USAID/IDRA survey is a good illustration of the political economy of governance indicators, broadly speaking. The questionnaire used for the survey was designed with the help of two global consulting firms (Casals and Associates and DPK Consulting). These firms made sure that the questionnaire and the presentation (formatting) of the final results aligns with the global language of corruption expertise. This omnipresence of global consultancies in the making of governance indicators speaks to a broader phenomenon of the privatisation of the state and an expansion of neoliberal governmentality (Barry et al. 1996; Miller and Rose 2008; Cordella and Willcocks 2010).

Besides questions of funding, this process also opens up an important discussion around the production and circulation of particular forms of knowledge and expertise. For instance, the core of the USAID/IDRA survey is generated through a global process of knowledge exchange. The original questionnaire for the survey was designed by the LAPOP (Latin American Public Opinion Project) opinion poll centre based in Vanderbilt University (USA), a centre that produces another important indicator, the Latin America Barometer, which is one of the sources of data for the World Bank's Worldwide Governance Indicator (WGI). The core questions regarding perceptions of corruption and the methodology for identifying and contacting respondents (the random route method) have remained the same in the 2005–2010 and the 2015–2016 surveys.

There are clear hierarchies of knowledge and funding that shape this field of global expertise and governance. A North-South dynamic is evident in the hiring of the American-based consultants whose job is to align the local knowledge to global and American forms of expertise and public relations. Such hierarchies reproduce historical relations of inequality among the global north and south and among specific kinds of expert knowledge. At the same time, it would be too simplistic to assume that these knowledge transfers are one-directional. Indeed, the research on IDRA's indicators over time also suggests that the centre's staff continuously engage in the creative labour of "translating" and repurposing corruption expertise to local scenarios and thus of incorporating local ways of knowing, evaluating, and judging corruption into the survey questionnaires. Rather than a one-directional flow of knowledge from North to South, these practices reveal a multi-directional process of

knowledge production. This process is indeed fraught with friction, constantly changing and often shaped by intended and unintended actors and events.

Measuring Perceptions, Experience, and Judgement

The USAID/IDRA questionnaire combines questions about perceptions, experiences, and assessments of corruption. The questions about perceptions take precedence in the final results of the survey. This set of questions is modelled on other corruption perception indicators (such as TI's CPI). One notable difference from TI's CPI is the pool of interviewees: while the CPI measures the perceptions of experts and the business elite, the USAID/IDRA survey collects data from the general public, public employees, and court staff. Despite these differences, what both indicators share is the production of a particular "form of knowledge" (Davis et al. 2012a, b) about corruption and governance that privileges aggregated data of perceptions of corruption.

IDRA's questionnaire also includes questions that seek to measure the experience of everyday corruption in interactions with public officials. This set of questions is also similar to other global indicators (such as TI's Bribe Payers' Index (BPI). Given the reluctance of respondents to admit to having paid a bribe, however, these questions remain under-reported, and hence not a good indicator of levels of corruption in the form of bribing.

Finally, the USAID/IDRA survey contains a set of questions that measure "attitudes towards corruption". These questions address culturally specific notions of corruption, justice, and impunity. These questions assess respondents' notions of legitimacy or illegitimacy of given transactions with or by public officials. Respondents are asked to assess whether a public official (a teacher, a minister, a civil servant) or private entity (an individual, a businessmen) is corrupt or honest, should be punished or not for accepting or offering a bribe in exchange for a public service (a birth certificate, a grade, a public-private contract) [see, for instance, IDRA 2008, Q AOC1-10]).

The scenarios conjured in these questions are designed around the most common practices deemed illegitimate or corrupt, from bribing a doctor or a teacher, to favouring business groups in public procurement contracts. During conversations and interviews with the author in 2008, the IDRA staff shared examples from personal experiences of scenarios similar to the ones described in the questionnaires. One staff member, for instance, shared his experience of discomfort and indignation when at a doctor's office he was pressured to give a bribe for a service that was covered by the state insurance. Others talked about the ubiquitous corruption in the courts and the impossibility of obtain-

ing justice on court rulings without bribing. Others still took it as a given that ministers took a cut from public procurement contracts. I found these stories of everyday corruption reflected in a number of IDRA's surveys. For instance, the "Impunity: Perceptions and Experience of Albanian Citizens, 2014–2015" (IDRA 2015) survey included a series of questions that presented respondents with real life situations (such as witnessing bribe-giving or favouritism in a public tender) and asked them to rate the likelihood that they would report the illegal transaction to the authorities (IDRA 2015, p. 31). It further asked responders to rate their perception of the likelihood that the illegal transaction would be punished or not by justice institutions in Albania.

A number of the "attitude" questions developed across these different surveys are closely shaped by IDRA's staff. The scenarios conjured in these questions emerge from vernacular experiences with everyday corruption. Though not generating the data that captures the headlines, these questions are critical as they allow for the articulation of different official and unofficial legal cultures. Further, the process through which these questions are designed points to the agency of local actors in the making of expertise about corruption and governance. That said, there is no doubt that IDRA's survey data gain political traction only through the interventions of international actors in the local public sphere.

Using Indicators to Denounce Corrupt Affairs

To study the effects of the USAID/IDRA survey, the research for this chapter followed the circulation of several editions of the survey from 2008 to 2016. The production and circulation of two such editions (IDRA and DPK Consulting 2008; IDRA 2009) was closely monitored. These editions were produced during a time of lively political debate that centred on several allegations of high-level corruption, locally referred to as "afera korruptive" (corrupt affairs). "Afera korruptive" refers to cases of "legal corruption" or "state capture" (Hellman and Kaufmann 2001). Typically, "afera korruptive" refers to public-private partnership deals or public procurement contracts that follow the letter of the law, yet, clearly benefit specific public officials or their associates. In public debate and investigation of two alleged "afera korruptive" during 2008–2009 and in debate and voting on the judicial reform in 2016, a number of actors continued to draw upon the USAID/IDRA survey as a source of hard evidence of corruption. In the following, this chapter will trace how the results of the survey were used by different local and international actors on the ground.

Two "afera korruptive" were investigated by Prosecutor General Ina Rama in an effort to prove the judiciary's independence and its commitment to fighting corruption at the highest levels. One involved the public procurement process for a public works contract to build the so-called "Patriotic Highway" (Rruga e Kombit), a road linking coastal Albania to the country's border with Kosovo. In 2008, the prosecutor's office pressed charges against Transport Minister Lulëzim Basha, alleging abuse of office and violation of procedures in the tender for the contract by the multinational corporation Bechtel-Enka. Rama claimed that the minister had accepted unit costs of construction material greatly above the going market rates, costing the public budget an estimated total of USD 337 million (Brunwasser 2015). A second case targeted Defence Minister Fatmir Mediu, who was investigated in relation to explosions in an army depot, revealing a suspicious public procurement process for the country's ammunition and a network of arms dealing involving an American subcontractor and breaching key provisions in American and NATO laws on international arms trading. These two cases were the first since the democratic transformations of the early 1990s to target sitting ministers and members of parliament. As such, they represented a test of Albania's judiciary at a time when the country fared poorly in corruption rankings. It was in this context that the USAID/IDRA surveys came to play an important role in the power struggles between various local and international actors.

Corrupt Affairs, Corruption Data, and Anti-corruption Laws

The 2008 IDRA/USAID survey results were announced days before a vote in parliament regarding the law on the immunity of members of parliament. The law under discussion would repeal existing provisions for immunity, making it easier for prosecutors to investigate high-level officials.

This parliamentary debate around the law on immunity developed alongside a heated public debate around the alleged Gërdec' "afera korruptive". Allegations of high-level corruption followed explosions at the munitions depot in the village of Gërdec, located in the vicinity of the capital Tirana. On 15 March 2008, a chain of explosions erupted in the depot killing 22 people and injuring 300 others (see also Chivers et al. 2008; Klosi 2009). The explosions, dubbed by Prime Minster Berisha a "technological accident", raised a number of questions regarding the privatisation of the country's communist era ammunition. More specifically, the explosion brought

scrutiny to the procurement process for the dismantling of a massive amount of ammunition that was public property inherited from the highly militarised communist regime of Enver Hoxha. This public procurement contract issued by the Ministry of Defense and approved by the Council of Ministers enabled the transfer of the disassembling process to a private entity without a proper tender process and evading established safety regulations and expert oversight. The private entity was a joint venture between the American company Southern Ammunition Company (SAC) and the Albanian firm, Albademil, owned by the infamous Mihal Delijorgji (an intermediary to the government company, Military Export Import Company, MEICO). Further, rather than being dismantled and destroyed, this ammunition was resold to US-backed Afghan troops via the American subcontractors, AEY, Inc.. As links to AEY Inc. were being investigated by the Pentagon, Albania came under international pressure to conduct its own investigations into the case. The investigations into AEY led to prison sentences for the owners of the firm, revealing a breach of the US laws on the sale of Chinese ammunition to American-backed troops. They also revealed a chain of intermediaries, of backroom deals, and of partnerships with corrupt public officials around the world willing to take a cut from arms sales' deals that breached local laws and norms (Chivers et al. 2008). The story of the two young guys from Miami who were able to set up AEY Inc. with a $300 million dollar contract from the Pentagon has recently been told in the book *War Dogs* (Lawson 2015; see also Snapjudgment 2015) and popularised in the movie by the same title. But the Albanian side of the story has received less attention (but see Klosi 2009).

Following the federal investigation by the Pentagon into AEY, Albanian prosecutor general Ina Rama began an investigation into the public procurement process of the Gërdec depot by the Albania's ministry of defence and MEICO. As more details about the sales of the dismantled ammunition became public, the Gërdec explosion was cast in public debate as a prime case of "afera korruptive".

Talk about "afera korruptive" has continued to dominate public debate in Albania and has implicated high-level officials across the political spectrum. This ongoing preoccupation with "afera korruptive" reflects, on the one hand, a global trend in anti-corruption discourse (namely, an increasing attention to legal corruption and state capture) and, on the other hand, a loss of trust in the judiciary (an established record of impunity of high-level officials). Alas, "afera korruptive" are harder to trace through the paper trail despite being deemed illegitimate by the broader public. As such they present a challenge to

perception- and experience-based indices of corruption. Because of the very design of their questionnaires and because of the nature of quantitative data on perceptions of corruption, specific cases of high-level corruption become lost in the generalised aggregated data of corruption indicators. It is only through various associations and circulations among various actors that such indicators regain political relevance.

For key local and international actors on the ground, prosecuting the alleged affairs of Gërdec and the Patriotic Highway became a test of the integrity and transparency of the judiciary—one of the branches of government with the poorest records for transparency and trust in institutions in previous USAID/IDRA surveys. In 2008, Ambassador Withers repeatedly appealed to local institutions for a fair investigation into the public procurement contracts and money flows between key actors involved in the Gërdec affair: MEICO, Edvin. Ltd, Albademil, and AEY Inc. Withers expressed his support for the prosecutor general Ina Rama and publicly criticised the parliament and the Prime Minister, repeatedly calling on them to refrain from influencing the investigations. In the Gërdec investigations, all eyes were on Fatmir Mediu, then minister of defence, and Ylli Pinari, director of MEICO. Mediu resigned from the post of minister immediately after the explosion but remained a member of parliament—which also granted him immunity from prosecution—and continued to hold his post as leader of the Republican Party—a strong ally of PD. In the Patriotic Highway investigations, the focus was on then minister of transport, Lulëzim Basha. Basha had requested the parliament to repeal his immunity, following an earlier request by the previous public prosecutor, Theodhori Sollaku (Likmeta 2008). Basha submitted himself to the investigations by the prosecutor general but escaped prosecution by contesting the formal case on procedural grounds. Basha continued to have a succesful political career. He is currently the leader of PD.

As the Gërdec and Patriotic Highway affairs unfolded in public debate, Ambassador Withers drew upon data from the USAID/IDRA survey to lobby the parliament to pass the law on the immunity of high level officials. This law would enable prosecutor Rama to fully investigate Mediu and Basha.

As the following section shows, international actors routinely use indicator data as a source of evidence to support specific policy and legal recommendations to local actors and institutions. These forms of intervention, however, have intended and unintended effects that may or may not contribute towards local anti-corruption efforts.

A Prosecutor, an Ambassador, and His Indicator

Ina Rama was appointed as Prosecutor General by President Bamir Topi and with the unfledgling support of Prime Minister and leading leader of PD, Sali Berisha. Rama was compared to the fictional character Silvia Conti, the uncompromising prosecutor general in the Italian television series *La Piovra*.[3] Berisha endorsed Rama precisely in an effort to break away from what he described as a corrupt judiciary inherited from the communist regime. Given mounting international pressure to grant the judiciary independence from political interests, Rama took to tackling corruption at the highest levels of government. In the period 2008–2011, she took three sitting ministers to court. Unfortunately, none of these cases resulted in convictions of high-level officials and both Rama's credibility as a prosecutor and as a person came under vicious attack from her former supporters. In 2011 she came near to being impeached by an ad-hoc Prime Ministerial commission investigating an alleged attempted coup.

Rama's rift with Berisha began precisely with the infamous Gërdec and Patriotic Highway cases. Despite overt pressure from Berisha to drop charges on Mediu and Basha, Rama continued her investigations. The Gërdec and Patriotic Highway cases presented Rama with a test of her public image as the uncorrupted prosecutor who would shift the widespread perception of corruption in the justice system. Throughout these investigations and the evolving conflict with Berisha, she had Ambassador Withers' full support. Speaking about Rama at various public events, Ambassador Withers invoked the analogy to Silvia Conti which had originally been introduced by Berisha (Likmeta 2012). By re-appropriating this analogy, Ambassador Withers invoked the image of the uncompromised young woman prosecutor who took on mafia penetration in the Italian justice system (see Musaraj 2015, pp. 235–236).

The Ambassador's involvement in the Gerdec had specific reasons relating to the American company involved in the affair. More broadly, however, this involvement has also been informed by data on corruption in Albania. Judicial institutions persistently score among the public institutions with the lowest levels of trust and transparency in Albania.[4] Hence, reforming the justice system and raising the number of prosecutions of high-level public officials has been at the forefront of the USAID Rule of Law Program in Albania and continues to be an issue of high priority for internationals and locals alike.

Indeed, judicial reform has returned to the main political agenda in Albania with a recent campaign to pass radical changes to the constitution and align

judicial institutions to EU norms. The reform, which entails the rewriting of several articles of the constitution and the reshuffling of existing judicial institutions, has once again become a major focus of international influence. Throughout 2016, EU and US ambassadors (Ambassador Romana Vlahutin and Ambassador Donald Lu, respectively) actively participated in the negotiations between the major political parties in Albania over the passing of the judicial reform. Lu has gone so far as to target individual politicians (namely, the current head of PD, Lulëzim Basha) and to issue explicit threats to members of parliament (such as blacklisting by the US government in case of a no-vote) (Erebara 2016). In summer 2016, thanks primarily to the international pressure, the reform passed with a consensus by both major political parties.

In addition to persistent involvement and effort to influence local political debate and policy-making in Albania, what we can observe in Ambassadors' Withers' and Ambassador Lu's interventions is a recourse to indicators as a source of objectivity and neutral data and as evidence of the vernacular experiences and perceptions of corruption. In addition to appeals to local images of incorruptibility, to boost Rama's legitimacy on the local political scene Ambassador Withers repeatedly referred to the USAID/IDRA survey as a true reflection of the integrity of the prosecutor. Coming to her defence at the time of the vicious campaign against her by Berisha's supporters, Ambassador Withers cited the next edition of the USAID/IDRA survey noting: "The good news is that the public perception of the public prosecutor has improved more than any other institution" (Withers, cited in *Tema* 2009, p. 7). For Withers, the survey provided objective evidence on the progress of the prosecutor general vis-à-vis other justice institutions. By referring to these data, the Ambassador used the indicator as a technology of governance, presenting its evidence as superior to other forms of knowledge (for instance, court records, public statements and so on).

Similarly, in 2016, Ambassador Lu appealed to the recent IDRA corruption survey to make a case for judicial reform in Albania. Speaking at the launch of the survey, Lu reflected on what a local politician had said to him: that Albanians do not really understand judicial reform. By contrast, noted Lu, the data from IDRA's survey shows that: "party leaders, prosecutors, judges, and customs officials have the biggest problems with corruption" (RTSH 2016). Further he noted that this did not surprise him; rather, this showed that "the Albanian people understands very well what is happening with corruption in Albania today. It is time for politicians to listen" (RTSH 2016). The last comment was a direct call for PD MPs to vote for the judicial reform.

This form of aggressive influence in local political decision-making has been consistent across the different US Ambassadors in Albania though the alliances with specific local political parties and actors has shifted over time.[5] Consistent is also the discursive strategy of citing corruption perception data (also sponsored by American institutions) as a source of evidence of corruption, evidence often framed as the voice of the people of Albania. At the same time, the effects of such uses of indicator data are not always anticipated.

While Withers used the indicator as a means to assert US influence over local political decision-making, Rama was an unexpected beneficiary of this political maneouver at a time when she was threatened by an authoritarian political system. In autumn 2008, Berisha's supporters in parliament proposed a law that would require the prosecutor general to respond directly to the parliament—a law that seriously challenged the separation of powers. Ambassador Withers' praise of the perceptions of the Prosecutor's office mentioned earlier came right before the debate in parliament over this law. The law did not pass. Withers' aggressive statements of support of the office of the prosecutor and of Rama's integrity may have played some role in the failure of this attempt to take over the judiciary.

Meanwhile, the effects of these interventions on the prosecution of the alleged "afera korruptive" have been modest if not insignificant. Despite Ambassador Withers' political pressure, the Gërdec case concluded with no charges against any high-level officials. Given the law on immunity from prosecution, the prosecutor's office was not able to investigate Fatmir Mediu, one of the key figures involved in the procurement process of the ammunitions factory in Gërdec. Other officials also escaped prosecution due to legal loopholes and administrative obstacles encountered throughout 2008–2009, the time-frame of the investigations by the prosecutor's office. Lulëzim Basha's case was also closed on procedural errors, avoiding a trial that would consider the charges against him. Further, it seems that these investigations did not even manage to prevent these actors' further ascent to power. Thus, Mediu continues to hold the office of leader of PR and Basha is leader of PD and contender for the post of prime minister (which he lost to Edi Rama in the recent elections of 2017). The third official not discussed in this chapter but also prosecuted (unsucessfully) by Ina Rama on high-level corruption charges is Ilir Meta, former leader of a minor parti, LSI. Meta also resigned at the time of the investigations (2011) but was cleared of the charges (at the time his party was in coalition with PD; in 2013, Meta, joined the coalition with PS). He was recently elected president of Albania.

Perhaps the only modest positive effect of these political interventions was Ina Rama's survival as prosecutor general without being sacked or persecuted

herself. Ina Rama continued to face threats and slander from the incumbent party, and concluded her term in 2012 without seeking reappointment. Although she was not successful in pursuing a high-level corruption case, Rama's survival in her office in the course of the Gërdec and Patriotic Highway investigations would seem highly improbable without the constant backing of Ambassador Withers. Overall, looking at how these "afera corruptive" and the broader discussions about corruption in Albania developed over time, one notices an ongoing involvement of international actors, a mobilisation of indicator data about corruption, but little evidence of effective impact of this involvement and of this data on curbing corruption at the government level.

Conclusion

International actors have repeatedly played an important role in legitimising or de-legitimising various political parties in Albania, often on opposite sides of the political spectrum. With this context in mind, the interventions by Ambassadors Withers and Lu are not exceptional. What this research has shown, however, is that corruption and governance indicators (such as IDRA's corruption perception survey) play a key role in justifying these interventions.

Overall, the specific case of Albania speaks to continuities in the networks of indicators' production and the increasing role of indicators as a source of expertise in political decision-making (Sampson 2010; Davis et al. 2012b; Bhuta 2012; Merry et al. 2015). This authority of indicators as a superior form of expertise derives from the complex political economy and ecology of funding the production of indicators and their "translation" to local phenomena on the ground. At the same time, this research shows that indicators are also used by different actors on the ground to forward different agendas, often unanticipated by either the funders or the producers of this data. The following concluding discussion points to some pressing issues regarding the politics of producing, funding, and deploying corruption indicators in local and transnational contexts.

First, given the increasing importance of corruption indicators as measures of governance, it remains critical to question the political ecology and economy of the production of such indicators. As this chapter has tried to show, country or cross-country indicators are produced with the help of various intermediaries that cut across boundaries between private and public, global and local institutions. Thus, the making of IDRA's questionnaire involved a government agency (the USAID), global private consulting companies, and a

local research centre. Further, the production of the USAID/IDRA survey is contingent on the availability of appropriate funding streams. Assuming that the USAID/IDRA survey had an important role as a diagnostic of the progress against corruption in Albania, its abrupt interruption in 2010—due to the end of its funding stream—raises questions about the continuity and durability of such indicators—an aspect crucial to the type of evidence that they provide. As such, the USAID/IDRA survey serves as a limiting case and points to a tension between the business models that enable the production of governance indicators and the specific features—continuity, consistency in methods and samplings, repetition—of the type of data they produce over the course of time.[6]

A second concern addressed in this chapter relates to the use of governance indicators by international actors to influence local policymaking and institutional reform. The chapter traced the circulation of several editions of the USAID/IDRA survey data—from 2008 to 2016—in relation to local anti-corruption movements and initiatives. Thus, although the USAID/IDRA survey aimed at raising awareness about corruption, when used by local and international actors in the context of the "afera korruptive" of Gërdec and the Patriotic Highway, the survey, to quote Debora Valentina Malito, became a means of "judg[ing] rather than monitor[ing …] the contemporary political order" (2014, p. 9). Furthermore, the recurring citations of the USAID/IDRA corruption survey data by international actors (in this case, Ambassadors Withers, Lu and Vlahutin) in key moments of local political and legal debate suggest that governance and corruption indicators often serve as a means of directing policy.

Finally, following the several editions of the USAID/IDRA survey over time alongside local debates and initiatives that target corruption enables a better view of shifting discourses and targets of corruption. In the case of Albania, over the past two decades, the target of anti-corruption discourse and initiatives has shifted from petty corruption (early 2000s) to state capture (2008–2010) to corruption in the judiciary (2016–2017). These moments have intersected with a global global shift of corruption discourse from petty corruption to state capture (late 2000s) and the intensification in the last decade of anti-corruption measures in the EU accession process. These intersecting histories of anti-corruption invite us to rethink the short- and long-term effects of corruption and governance indicators in local political contexts. In the short term, corruption indicator data seems to have an effect on local political debates, especially when used by international actors such as US ambassadors or EU representatives. In the course of the unfolding of the Gërdec case, for instance, the survey data corroborated allegations of state capture in a public procurement process conducted by the Ministry of Defence. While not effective at building a case against responsible high-level officials, it is arguable that, indirectly, the use of this indicator data by

Ambassador Withers helped shield the prosecutor general from executive pressure (and from possible prosecution). Likewise, the indicator data was recently used by Ambassador Lu to make a case in support of the judicial reform passed in 2016. In both cases, the indicator data was read as the true measure of corruption in contemporary Albania.

Ina Rama's career (and her image of the Albanian Silvia Conti) would seem to come full circle with the passing of the judicial reform. Unable to find a job in the judiciary since the end of her tenure as prosecutor general, Rama was elected as one of the seven judges in the newly established Appeals Council. This is an independent body that is at the core of the judicial reform. It is mandated to monitor other judicial branches to prevent and combat corruption in the judiciary. It remains to be seen if Rama's second return will indeed deliver the anti-corruption promisses of the judicial reform. In the meantime, what we learn from these intersecting stories of indicators, ambassadors, prosecutors and anti-corruption initiatives is that the effects on the ground of governance and corruption indicators are subject to the specific configurations of local political cultures and networks of political actors.

Acknowledgements Research for this project received the support of the National Science Foundation and the Social Sciences Research Council. The ethnographic research for this chapter has benefited from an internship with IDRA and interviews with various staff members during 2008. The author thanks in particular the director of IDRA, Auron Pasha, for providing access to IDRA's resources and to the IDRA staff for their time and support during research conducted in 2008–2009. The opinions and analysis expressed in this chapter do not reflect those of IDRA staff, nor of the USAID. Earlier drafts of the chapter were circulated in the Indicators and Global Governance research network workshops and the Global Governance by Indicators workshop. The author thanks in particular Sally Engle Merry, Kevin Davis, Nehal Bhuta, Debora Valentina Malito, Gaby Umbach, and other participants in these networks for their comments and suggestions. Last but not least, special thanks are due to Rozeta Koçi and Miranda McKinnon for research assistance.

Notes

1. The indicator in question acts as a "quasi-indicator" (Davis et al. 2012b). It compares institutions to one another and over time. However, unlike other indicators (such as TI's CPI, for instance), its rankings are not compared to those of other countries.
2. It should be noted that the local branch that managed the Albania scores for the TI CPI, the Citizens Advocacy Office (CAO), also received funding support from the USAID Rule of Law Program.
3. As detailed in Musaraj 2015, the Italian TV mini-series *La Piovra*, produced in Italy between 1984 and 1999, was broadcast on the highly censored (only)

Albanian television station between 1987 and 1988. The TV series was extremely popular in a number of late socialist countries (among others, the former USSR, Bulgaria, and Albania). The series depicts the Italian mafia and its multiple tentacles entwined around the highest levels of government—hence the metaphor of the octopus. Silvia Conti, a deputy prosecutor, is one of the few characters in the series who is not corrupt.

4. According to the USAID/IDRA survey, the judiciary remained one of the least trusted institutions throughout 2005–2009 (IDRA 2010, p. 14).
5. A recent account of these transformations highlights the enormous support that American institutions provided for Berisha and the PD in the beginnings of the democracy movements (Abrahams 2015). Since the Gerdec investigations, the support of the American Embassy has shifted towards the Socialist Party.
6. Similar questions of continuity in the production of data also emerge from internal debates about the best methodological tools to use. This is, for instance, an issue that also applies to the recent changes in TI's CPI, which has changed it methodology making it impossible to compare the data from the 2011 CPI with that of the 2012 CPI (TI 2012).

References

Abrahams, F. (2015). *Modern Albania: From Dictatorship to Democracy in Europe*. New York: New York University Press.

Barry, A., Osborne, T., & Rose, N. (Eds.). (1996). *Foucault and Political Reason: Liberalism, Neo-liberalism and Rationalities of Government*. Chicago: University of Chicago Press.

Bhuta, N. (2012). Governmentalizing Sovereignty: Indexes of State Fragility and the Calculability of Political Order. In K. Davis, A. Fisher, B. Kingsbury, & S. E. Merry (Eds.), *Governance by Indicators: Global Power Through Quantification and Rankings*. Oxford: Oxford University Press.

Brunwasser, M. (2015, January 30). *Steamrolled: A Special Investigation into Business Abroad. Foreign Policy*. Retrieved from http://foreignpolicy.com/2015/01/30/steamrolled-investigation-bechtel-highway-business-kosovo/

Chivers, C. J., Schmitt, E., & Wood, N. (2008, March 27). Supplier Under Scrutiny on Arms for Afghans. *The New York Times*, p. A1.

Cordella, A., & Willcocks, L. (2010). Outsourcing, Bureaucracy and Public Value: Reappraising the Notion of the "Contract State". *Government Information Quarterly, 27*(1), 82–88.

Davis, K., Fisher, A., Kingsbury, B., & Merry, S. E. (Eds.). (2012a). *Governance by Indicators: Global Power Through Quantification and Rankings*. Oxford: Oxford University Press.

Davis, K., Kingsbury, B., & Merry, S. E. (2012b). Indicators as a Technology of Global Governance. *Law and Society Review, 46*(1), 71–104.

Dutta, N. (2015). Tradeoffs in Accountability: Conditionality Processes in the European Union and Millennium Challenge Corporation'. In S. E. Merry et al.

(Eds.), *The Quiet Power of Indicators: Measuring Governance, Corruption, and Rule of Law* (pp. 156–196). Cambridge: Cambridge University Press.

Erebara, G. (2016, May 5). US, EU, Warn Albania Against Judicial Reform Delay. *Balkan Insight*. Retrieved from http://www.balkaninsight.com/en/article/us-and-eu-ambassadors-warn-albania-leaders-about-much-delayed-justice-reform-05-05-2016-6

Gupta, A. (1995). Blurred Boundaries: The Discourse of Corruption, the Culture of Politics, and the Imagined State. *American Ethnologist, 22*(2), 375–402.

Halliday, T. (2012). Legal Yardsticks: International Financial Institutions as Diagnosticians and Designers of the Laws of Nations. In K. Davis, A. Fisher, B. Kingsbury, & S. E. Merry (Eds.), *Governance by Indicators: Global Power Through Quantification and Rankings*. Oxford: Oxford University Press.

Hellman, J., & Kaufmann, D. (2001). Confronting the Challenge of State Capture in Transition Economies. *Finance and Development, 38*(3), 31–35.

IDRA & DPK Consulting. (2008). *Corruption in Albania: Summary of Findings*. Tirana: IDRA & USAID.

IDRA & FSHSH. (2015). *Impunity: Perceptions and Experiences of Albanian Citizens, 2015*. Tirana: IDRA & FSHSH. Retrieved from http://www.idrainstitute.org/files/reports/impunity_2014/fact_sheet_sq.pdf

Institute for Development Research and Alternatives (IDRA). (2009). *Corruption in Albania: Summary of Findings*. Tirana: IDRA & USAID.

Institute for Development Research and Alternatives (IDRA). (2010). *Corruption in Albania: Summary of Findings. Survey 2010*. Tirana: IDRA & US Agency of International Development (USAID).

Institute for Development Research and Alternatives (IDRA). (2016a). *Corruption in Albania: Perceptions and Experience 2015–2016*. Tirana: IDRA. Retrieved from http://www.idrainstitute.org/files/reports/Corruption%202016/Corruption%20English%20FINAL.pdf

IDRA. (2016b). *Judicial Reform in Albania: Knowledge, Support and Expectations*. Tirana: IDRA. Retrieved from http://www.idrainstitute.org/files/reports/judicial_reform_2016/Graphs_survey.pdf

Kajsiu, B. (2013). The Birth of Corruption and the Politics of Anti-Corruption in Albania, 1991–2005. *Nationalities Papers: The Journal of Nationalism and Ethnicity, 41*(3), 1–18.

Kajsiu, B. (2016). *A Discourse Analysis of Corruption: Instituting Neoliberalism Against Corruption in Albania, 1998-2005*. London: Routledge Press.

Klosi, A. (2009). *Katastrofa e Gërdecit: Shkaqet, Shkaktarët, Viktimat*. Tirana: K&B.

Lawson, G. (2015). *War Dogs*. New York: Pocket books.

Likmeta, B. (2008, December 22). Albania's Patriotic Highway' Scandal. *Bloomberg News*. Retrieved from http://www.bloomberg.com/news/articles/2008-12-22/albanias-patriotic-highway-scandalbusinessweek-business-news-stock-market-and-financial-advice

Likmeta, B. (2012, November). Saying Farewell to an Albanian Hero in the Fight Against Corruption. *Foreign Policy*. Retrieved from http://transitions.beta.dev.foreignpolicy.com/posts/2012/11/16/saying_farewell_to_an_albanian_hero_in_the_fight_against_corruption

Malito, D. V. (2014). *Measuring Corruption Indicators and Indices* (Robert Schuman Centre for Advanced Studies Research Paper 2014/13). Florence: European University Institute. Retrieved from http://dx.doi.org/10.2139/ssrn.2393335

Merry, S. E. (2016). *The Seductions of Quantification: Measuring Human Rights, Gender Violence, and Sex Trafficking*. Chicago: University of Chicago Press.

Merry, S. E., Davis, K., Fisher, A., & Kingsbury, B. (Eds.). (2015). *The Quiet Power of Indicators: Measuring Governance, Corruption, and Rule of Law*. Cambridge: Cambridge University Press.

Miller, P., & Rose, N. (2008). *Governing the Present: Administering Economic, Social and Personal Life*. Cambridge: Polity Press.

Musaraj, S. (2011). Tales from Albarado: The Materiality of Pyramid Schemes in Postsocialist Albania. *Cultural Anthropology, 26*(1), 84–110.

Musaraj, S. (2015). Indicators, Global Expertise, and a Local Political Drama: Producing and Deploying Corruption Perception Surveys in Albania. In S. E. Merry, K. Davis, A. Fisher, & B. Kingsbury (Eds.), *The Quiet Power of Indicators: Measuring Governance, Corruption and Rule of Law* (pp. 222–247). Cambridge: Cambridge University Press.

Rottenburg, R., Merry, S. E., Park, S. J., & Mugler, J. (Eds.). (2015). *The World of Indicators. The Making of Governmental Knowledge Through Quantification*. Cambridge: Cambridge University Press.

RTSH. (2016, June 2). *Lu: Politikanet te degjojne popullin*. Retrieved from http://rtsh.al/lajme/lu-politikanet-te-degjojne-popullin/

Sampson, S. (2005). Integrity Warriors: Global Morality and the Anticorruption Movement in the Balkans. In D. Haller & C. Shore (Eds.), *Corruption: Anthropological Perspectives* (pp. 103–130). London: Pluto Press.

Sampson, S. (2010). The Anti-Corruption Industry: From Movement to Institution. *Global Crime, 11*(2), 261–178.

Seligson, M., & Baviskar, S. (2006). *Corruption in Albania: Comparisons Between 2004 and 2005 [Survey Data Set]*. Retrieved from http://www.vanderbilt.edu/lapop/albania/GoodGovernanceinAlbaniav82r.pdf

Serban, M. (2015). Rule of Law Indicators as a Technology of Power in Romania. In S. E. Merry et al. (Eds.), *The Quiet Power of Indicators: Measuring Governance, Corruption and Rule of Law*. Cambridge: Cambridge University Press.

Shekulli. (2008, May 8). Withers: Fatkeqesisht 92% e qytetareve pranojne korrupsionin (Unfortunately 92 % of Albanians agree that there is corruption).

Snapjudgment. (2015). *The Accidental Arms Dealer. Snapjudgement, Storytelling with a Beat*. Retrieved from http://snapjudgment.org/accidental-arms-dealer

Tema. (2009, May 7). *Withers: Politika nuk duhet te ushtroje presion ndaj qeverise* (Politics Should not Exert Pressure on the Government).

Transparency International. (2012). *Corruption Perceptions Index 2012: Technical Methodology Note.* Berlin: Transparency International. Retrieved from http://tiukraine.org/files/u/124/docs/g._technical_methodology_note_corruption_perceptions_index_2012_0.pdf

16

Activism Through Numbers? The Corruption Perception Index and the Use of Indicators by Civil Society Organisations

René Urueña

Introduction

As indicators are increasingly used as tools of global governance (see Kingsbury et al. 2012), one interesting dimension of this new reality is that civil society organisations are also using indicators in their work, as means for activism, contestation and reform. This chapter explores this dimension of indicators in the context of the global anti-corruption movement. In the most traditional scenario, civil society organisations use indicators (e.g. human rights rankings) in order to shame their respective governments into action. However, other less traditional modes of engagement also exist. Sometimes, civil society organisations create indicators to give visibility to an issue they find relevant. For example, research on the use of indicators in the context of the Colombian armed conflict reveals a civil society organisation (Human Rights and Displacement Consultancy, CODHES) that does activism by producing indicators on internally displaced people and then confronts the national government with the figures (Urueña 2012a, b). In other contexts, indicators are used by civil society groups as tools for planning and evaluating humanitarian responses, such as in Haiti (Satterthwaite 2010; Satterthwaite and Moses 2012). As a result of such use, sometimes the production of indicators by intergovernmental bodies ends up empowering local NGOs. For example, it is often the case that international institutions lack the capacity to gather the

R. Urueña (✉)
Universidad de Los Andes, Bogotá, Colombia

D.V. Malito et al. (eds.), *The Palgrave Handbook of Indicators in Global Governance*,
https://doi.org/10.1007/978-3-319-62707-6_16

raw data required for indicators and they have to either rely on self-reporting by states or else on information from NGOs operating on the ground (Kälin and Künzli 2009). This means that assessment of compliance by, say, a human rights body is often just a reflection of data provided by the state whose compliance is being evaluated, or by NGOs with an agenda of their own (Alston 2000).

How is civil society activism affected by this penetration of quantitative technologies of governance? This chapter explores such questions by focusing on one of the best-known civil society efforts at quantification: the Corruption Perception Index (CPI) developed and published by Transparency International. It argues that the use of indicators opens new avenues of activism and persuasion for civil society organisations, while at the same time imposing important costs that may, on occasion, balance the scales against their use. Indicators, moreover, seem to be a tool that is neither of use nor available to all civil society organisations. By studying the case of Transparency International as an organisation, this chapter shows that this organisation's particular characteristics (technically sophisticated, socially networked, emerging from a global professional elite) may explain more about the success of the CPI than any characteristic of the Index in itself. Finally, the chapter explores the role played by the Index in the wider landscape of framing the problem of corruption as a structural challenge to economic development, and its role both as a prism through which to observe reality and as an artefact to constitute the very reality that it observes.

The chapter proceeds in the following way: the first section introduces Transparency International and the ideological and institutional context in which it first emerged. The next two sections explore the role of the CPI and the difficult trade-offs it demanded from Transparency International as a civil society organisation. The last two sections discuss the role of the Index in constituting the "reality" underlying the anti-corruption movement and propose some conclusions.

The Context of Transparency International

The rise of corruption indicators follows a similar trajectory to that of other governance indicators. As Katharina Pistor has shown, governance indicators are hardly something new. Ever since the early 1970s, private firms have compiled governance indicators in order to provide business decision-makers with tools to assess risk. For example, the Business Environment Risk Intelligence ratings (BERI) were first compiled in 1972, and the International Country

Risk Guide (ICRG) in 1980 (Pistor 2012). However, a crucial breakthrough came in 1995 with the publication of the first Corruption Perception Index, developed by Transparency International (TI). The present section explores the story behind this indicator and its connections to TI's wider ideological, institutional and epistemological context. Ultimately, the section argues that the CPI emerged in an ideological context that understood corruption as a key obstacle to economic development, it rose from the institutional milieu that produced that mindset, and it serves the epistemological needs of the global anti-corruption campaign.

TI is an international non-governmental organisation whose stated mission is the "relief of poverty, suffering and distress in any part of the world caused directly or indirectly by corruption" (Transparency International 2016a). TI's governance structure has two key components: the first is the Secretariat, based in Berlin, which is accountable to the International Board and to the wider TI movement through the Board of Directors and the Annual Membership Meeting. The Secretariat is led by an Executive Director (currently the South African Cobus de Swardt) and it had an operating budget of about 25 million euros in 2014 (Transparency International 2016b). The second key component is constituted by the national chapters, which are local non-governmental organisations that have been accredited by the Board of Directors under the National Chapter Accreditation and Individual Member Appointment Policy, which imposes a set of procedures and requirements for the accreditation process. There are currently more than 100 national chapters, which vary in size, expertise and access to funding.

TI was founded in 1993 by Peter Eigen, a former senior World Bank official, together with other individuals with extensive international experience, such as Fritz Heimann, a legal counsel from General Electric, and Frank Vogl, also a former official of the World Bank. Despite being created by former World Bank staff, TI is not a spin-off of World Bank policies. On the contrary, according to his own account, Eigen created TI because the World Bank was unwilling to push further in understanding corruption as a crucial part of its mandate (Lundberg 2002). The Bank's position was that corruption was a political issue, and that the Bank should not get involved in such debates. In fact, the Bank's initial approach was to deny funding to TI and to block efforts to create an "island of integrity" (Lundberg 2002), one of TI's initial initiatives, around Bank-funded projects. Despite this initial difference of views, TI's approach seems to have set a path that was then followed by the World Bank. At the 1996 Annual General Meeting, the Bank's President James Wolfensohn referred to corruption as a "cancer" (Wolfensohn 2005), thus opening the way for a more explicit discussion of the subject within the Bank

(Fjeldstad et al. 2008, p. 2). It is interesting to note, though, that in the three-year period between TI's launch and the World Bank's shift, the staff at the Bank saw TI as a vehicle to push the issue of corruption while maintaining enough distance to afford deniability despite the World Bank Board's reluctance to engage with it (Fjeldstad et al. 2008, p. 2).

Despite these differences in political strategy, it seems clear that TI's overall framing of corruption was part of the same mindset that prevailed in the World Bank. The issue was, in essence, a matter of timing: due to political and institutional inertia, the World Bank required more time to catch up with the anti-corruption campaign that its former staff were already leading. It then caught up, and the overall landscape seems to be one of general agreement between the anti-corruption movement (as represented by TI) and the Bank.

What was this general mindset? The rise of the current global anti-corruption movement may be explored as a result of at least two interconnected factors: the first is the adoption of the United States Foreign Corrupt Practices Act of 1977 (FCPA), which criminalised the corruption of foreign officials by Americans. The second is the crystallisation of the link between corruption and economic development which became part of the common sense in academic and multilateral financial institutions in the 1990s.

The FCPA is an example of a first-mover situation. The US was risking the competitive advantage of its economy and business sector if other rich economies failed to adopt similar domestic legislation in turn criminalising their own nationals for corruption abroad. To prevent this outcome, the US engaged in an active diplomatic effort to create an even playing field, particularly by encouraging the adoption of multilateral anti-corruption instruments (Razzano and Nelson 2008). By any measure, this was an extremely successful undertaking: from 1997 to 2007 at least seven major multilateral instruments were adopted. First and foremost was the 2003 United Nations Convention against Corruption (UNCAC), which—along with other regional and multilateral instruments that were adopted in the 1990s and 2000s[1]—represented a crucial victory for the global anti-corruption movement. In turn, this network of international instruments triggered the adoption of multiple domestic statutes, some of which go further than the FCPA. Countries like the UK and Germany, for example, have become prominent in their prosecution of foreign corrupt practices, thus creating a network of anti-corruption efforts that gives the campaign a truly global dimension.

The political dynamics triggered by the FCPA were mostly focused on the developed world. Capital-importing countries did have their own domestic laws that criminalised corruption, but the challenge was that they did not

seem to be operative: local anti-corruption institutions seemed weak in the face of pressure from deep-pocketed foreign investors and their local political allies and, even if the local authorities did take action, foreign investors could still escape prosecution due to jurisdictional limitations. The answer was, therefore, to focus on action to be taken by capital-exporting countries and to create the legal obligation to prosecute there.

This approach was complemented by the adoption of anti-corruption measures in developing countries. In this case, however, the key dynamic was a transformation in the common sense of international financial institutions with regard to corruption. As is well known, the 1980s brought a new kind of developmental thinking that favoured liberalisation and market-based solutions[2] (Williamson 2005). The Washington Consensus came up with its very own anti-corruption agenda, and insisted on the risks of rent seeking. According to this view, state intervention in the market created opportunities for public officials to use their position to gain wealth. Hence, total deregulation and liberalisation would prevent this risk. Ultimately, in the most orthodox neoliberal view, the smaller the state, the smaller the risk of corruption. However, by the early 1990s the idea of complete liberalisation came under criticism. Instead, the idea that "institutions matter" for development gained momentum. The basic idea behind this new paradigm was that complete deregulation was improbable and perhaps undesirable, so the challenge was to get the institutions right. In this context, institutions became crucial, as they established the rules of the game for economic development. In the words of a well-known proponent at the time, "the standard constraints of economics (…) define the choice set and therefore determine transaction and production costs and hence the profitability and feasibility of engaging in economic activity" (North 1991, p. 97; see North 1990 for a general overview).

If institutions were crucial for economic activity their failure would undermine economic development, and a central institutional failure is, of course, corruption. Following this rationale, the 1990s saw the emergence of a robust body of literature that correlated corruption with slow economic growth (Shleifer and Vishny 1993; Mauro 1995, 1996). Political reality also underscored this link: massive deregulation and privatisation in Latin America and the former Soviet Union created new opportunities for corruption, which became notorious in many transitional economies. This intellectual and political shift put anti-corruption squarely in the mandate of international financial institutions, particularly the World Bank (Polzer 2001).[3] Since the 1990s, corruption and good governance have become central to the agendas of multilateral financial institutions,

which have made the anti-corruption campaign an important aspect of the overall transformation of programme governance championed since then.

A crucial intellectual link in this mindset was that there was a connection between the hurdles that entrepreneurs had to face in order to establish a business and economic development. Peruvian economist Hernando de Soto famously argued at the time that red tape in the South led people to informality, and lack of proper title and enforceable contracts led to underdevelopment (de Soto 1989). The dominant paradigm of the day was that "institutions mattered"; hence, development could be achieved through institutional reform. This approach became influential in multilateral financial institutions, and should be read together with their turn towards anti-corruption in the development agenda. While complete deregulation was implausible, making it crucial to correctly design institutions, the rationale seems to have been that the more regulation business was subject to, the more corruption was expected to emerge. More lately, however, the focus on corruption has been tied to the more general idea of the rule of law. While the shift is not sharp, and corruption is still read instrumentally as an obstacle to development, a new view seems to be emerging according to which legal institutions are part of development themselves. In this sense, absence of corruption as part of the rule of law is a developmental goal in its own sake (Santos and Trubek 2006, pp. 6–9).

Niche and Strategy

Transparency International is part of a wider institutional complex which has a particular ideological dimension, in the sense that it mobilises meaning in order to affect the distribution of resources in the direction it favours. In our case, the anti-corruption campaign shares the ideological bent of a specific set of policies to foster development that have been implemented since the late 1980s, and shares an institutional milieu with the producers of these policies, particularly multilateral financial institutions such as the World Bank and regional development banks.

Despite these commonalities in approach and mindset, TI had to find its place in the anti-corruption movement's global landscape. On the one hand, its role could not be the same as that of a multilateral financial institution since it lacked the financial resources and governance structure required to have an impact playing that role. But it was certainly not a grass-root civil society organisation: since its beginnings, TI's niche was not denouncing

individual acts of corruption or campaigning for the prosecution of specific civil servants. As one of the founding members told an evaluation team from the Norwegian Agency for Development Cooperation, the founders "never saw [themselves] as a 'placard-wielding NGO'" (Norwegian Agency for Development Cooperation 2011, p. 7). Instead, TI focused on working with institutions from the inside, with a heavy emphasis on technical top-down solutions to corruption, building on the extensive network of professional contacts established by its founders.

In a similar line, TI was particularly media savvy from its early days. Perhaps due to the fact that its initial vice-chair person, Frank Vogl, had been a spokesman for the World Bank, TI had very early access to important media outlets. Its launch event was covered by *The Economist* (1993) and the *Financial Times* (Hollman 1993), and as early as October 1993, just eight months into its existence, it was the subject of a favourable profile in *The New York Times* (McNickle 1993). TI's particular anti-corruption niche was in this sense a mix between top technical and professional expertise generally in tune with the views of senior staff in multilateral financial institutions, an expanding network of semi-independent national chapters that were, however, expressly prohibited from undertaking individual investigations of corruption, and a network of contacts with high-level personalities from the media, finance and politics who provided both entry to global media outlets and the visibility and trustworthiness that local or grass-root anti-corruption movements can lack.

What work could an organisation with these characteristics do? TI's first approach was to focus on three fronts. First, it developed analytical tools in order to assess anti-corruption capacity and institutions in particular countries. They called this the "national integrity system", and it consisted in a guidebook of transparency best practices which then grew to become TI's Sourcebook on National Integrity Systems, first published in 1996. This was to be implemented with the cooperation of interested governments. In parallel, and this is the second front, TI started to work on a more short-term strategy: "islands of integrity", which encouraged private actors in a specific sector or deal to simultaneously sign anti-bribery pledges. Ecuador was the first pilot project, in 1994. Finally, a third area of work that quickly became interesting for TI was multilateral conventions against corruption: the organisation was active in lobbying the OECD to create an anti-corruption instrument. This would eventually become the 1997 OECD Convention on Combating Bribery of Foreign Public Officials in International Business Transactions. In the American continent, TI also influenced putting corruption on the agenda of the 1994 Summit of the Americas, particularly through

its then advisory board chair, Alberto Dahik Garzozi, a Princeton-trained economist who was also Ecuador's Vice-President. This effort led to the 1996 Inter-American Convention Against Corruption.

None of these efforts, however, gave TI distinctive visibility as an organisation. This would soon change with the work of Johan Graf Lambsdorff, a young PhD from Göttingen University in Germany, who did voluntary work for TI in 1994. Lambsdorff was working on an aggregation of polls on corruption perception, and called his project a "corruption perceptions index". His work did not imply new research, but rather an effort to standardise the available indicators of corruption perception. At first, TI's Peter Eigen was sceptical. As he reported to Kirsten Lundberg, he thought, "this is too much, too fancy, and much too theoretical, and forget it. I didn't support him very much" (Lundberg 2002, p. 18). Moreover, there were serious doubts about the robustness of the Index: peer review was lacking, and the Index had not been discussed within the organisation. Unlike national integrity systems, the islands of integrity, or lobbying for an international instrument, the CPI was not a strategic priority for TI (ibid.). TI's hand was forced, however, when the German magazine "Der Spiegel" obtained (as background) Lambsdorff's experimental index. Faced with a fait accompli, TI's Secretariat had to decide whether to acknowledge the project or to disown it. The Secretariat decided to acknowledge it. The decision, as we will now see, was far from obvious, and provides useful lessons for civil society organisations who are interested in adding indicators to their activism toolkit.

The CPI: Civil Society Organisations and Indicators as a Platform for Mobilisation

TI is particularly well placed to use a tool such as an indicator: its self-image is one of an organisation focused on technical expertise, whose interventions are not grass-roots but rather make use of high-level contacts in finance and international organisations. With this caveat in mind, this section argues that the choice faced by TI does reflect some of the promises and challenges that the use of indicators poses to civil society organisations. On the one hand, the CPI is a relatively cheap tool (no original research is required) that is guaranteed to give TI visibility in the media—particularly because the organisation already had access to key players and could easily position a new idea that was clearly of interest to journalists (it was, after all, the media that were asking questions about the Index to begin with).

Moreover, the CPI could help position the TI "brand". This is important, as TI is keenly aware of the importance of the integrity of its name and the international reputation it entails (Norwegian Agency for Development Cooperation 2011, p. 18). While this is generally true of numerous international NGOs, in TI's case this challenge was even more pressing. It should be recalled that TI's national chapters receive no funding from the International Secretariat. This means that, ultimately, the whole process of accreditation of national chapters is based upon the added value of the prestige and access that the name TI entails. If the CPI helps increase the reputation of the TI name, the value of the TI label for national chapters increases as well.

This seems to be case. The CPI became TI's best-known product, and it served as a vehicle to improve TI's visibility as an organisation. As is clear from Fig. 16.1, the number of times the expression "Transparency International" was referred to in written documents in the English language sharply increased after 1995 in clear correlation with the increasing number of references to the "corruption perceptions index".[4]

If covered by the media, the CPI also provided a platform for putting the spotlight on countries that TI felt were not doing enough to encourage transparency. While it is clear that the rankings in the CPI are not connected with the implementation of other TI initiatives, it is also true that the Index gave the organisation a different kind of tool which could embarrass countries that were badly ranked. If the Index proved reliable enough, and a certain country's low score was publicised enough, then that country would implement anti-corruption measures (in the form of other TI initiatives, or otherwise) which could better that country's score.

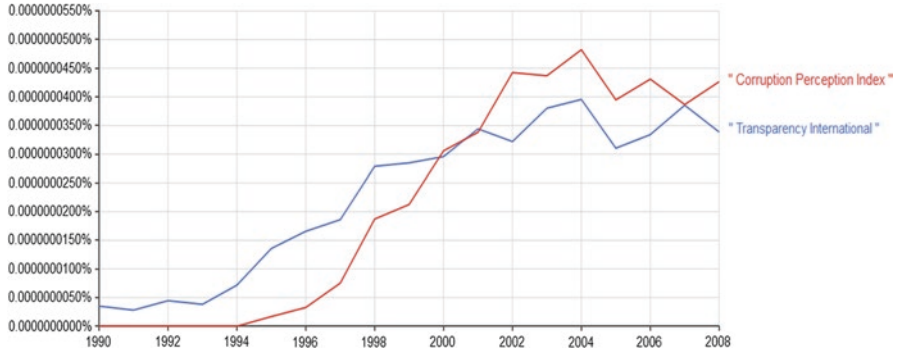

Fig. 16.1 Reference to "Transparency International" in online documents. Source: Own compilation based on Google n-gram

It is hard to assess whether this process has actually taken place. However, Arndt and Oman do report that, according to press coverage, the CPI has had an impact on the policy of governments as diverse as those of Cameroon, Papua New Guinea and Bosnia Herzegovina. The Republic of Korea, in addition, pledged to better its ranking in a certain number of years (Oman and Arndt 2006, p. 48). The CPI, and by extension TI, gains particular relevance when the Index is embedded in a larger process of governance. TI then plays the role of gatekeeper, administrating a valuable asset (CPI scores) that actors need for a further purpose. Such is the case of accession processes to the European Union, where good governance by aspiring members plays a central role. Heywood reports that east European countries (particularly Bulgaria, the Czech Republic, Poland, Romania and the Baltic States) took their score in the CPI very seriously, as they saw that a low ranking would deal a blow to their aspirations to join the EU (Andersson and Heywood 2009, p. 759). The relevance of the CPI can be similarly seen in the case of aid for development. For instance, USAID has used CPI scores to assess aid recipient's commitment to good governance and transparency (Andersson and Heywood, p. 758). This reality has been acknowledged by TI, which has noted that "some governments have sought to use corruption scores to determine which countries/territories receive aid, and which do not". TI discourages this practice, as "countries/territories that are perceived as very corrupt cannot be written off. Rather[,] they need help to emerge from the corruption-poverty spiral" (Transparency International 2009). While TI's position seems reasonable, the very fact that it is in the position to make this statement reveals the influence the CPI allows it to wield.

At the same time, however, the CPI has imposed important costs on TI. Perhaps the most important challenge has to do with the CPI methodology: ever since Lambsdorff's first pilot, it was clear that the methodology of the Index was open to debate and was going to be a target of severe criticism. It is beyond the scope of this chapter to open the issue of the CPI's methodology. Suffice it to say that the CPI triggers two different kinds of debates. The first is less concerned with the Index's methodology than with its effects: while TI is keen on emphasising that the Index is based on perception, and thus should not be taken to reflect "reality", the fact is that media coverage and the political impact of the CPI is partly based on journalists and policymakers actually taking the Index to reflect some kind of reality of corruption. In this sense, the Index may be seen to trigger effects that are not justified by its methodology (Cobham 2013). Moreover, other criticisms are indeed internal to the methodology: its definition of corruption, the challenges that are posed by the fact that the index is an aggregation of other indicators, the way in

which such scores are weighted and calculated, and the Index's apparent "perception lag", which fails to register that countries have improved their governance, among many others (see Sampford et al. 2006, chapters 2–6). What is important for our purposes, however, is that such critiques are not only directed at the CPI: they involve TI, and its political standing. TI's reputation is therefore compromised when the CPI is questioned.

Moreover, the CPI could also endanger TI's relations with its national chapters: while the Index is published by the Secretariat in Germany, the national chapters of countries that have bad scores could suffer retaliation from their own governments, who may hold them responsible as part of TI. This seems to be the case of Bangladesh, where government officials do not accept the local chapter's explanation that it has no responsibility for its country's CPI rating (Norwegian Agency for Development Cooperation 2011, p. 31).

Indicators and the Anti-Corruption Agenda: A Normative, Universal, and Facts-Based Style of Discourse

Beyond these opportunities and challenges, which could be deemed strictly practical, there is one further dimension to the use of indicators by civil society organisations (in this case, TI) which is worth noting. Earlier on, this chapter situated TI's emergence in its ideological and institutional context. Situating the organisation in the general landscape of the shifting ideas on law and development allows us to think of the technologies of its anti-corruption campaign in a different fashion—particularly concerning indicators such as the CPI. This is particularly useful in our case because the anti-corruption discourse often has a normative and globalising tone. Corruption is "wrong", and it is wrong everywhere: despite cultural differences, or the North/South and West/East divides, this is something everyone everywhere allegedly agrees on. Nevertheless, if the anti-corruption agenda is to be something more than just an empty call to do the "right thing", then it becomes enmeshed in debates about the model of development it champions.

These institutional and ideological dimensions are obscured by the tone of the anti-corruption campaign. However, they remain observable as the struggle against corruption often entails legal and institutional reforms (such as improved transparency, duty of disclosure and procedural rights of participation (Open Contracting Principles 2016) that have similarities to reforms that

seek to open developing economies to foreign investment, to increase the role of private power in the economy, and to diminish the role of public power in the redistribution of resources.

To be clear, it is not that transparency and participation are undesirable. The point here is not to criticise the underlying anti-corruption policies, which are often desirable in their own right (see Urueña 2012a, b), but rather to engage with them as expressions of a consensus that has ideological substance. To do this, it is important to acknowledge that the anti-corruption campaign is not aiming to make an intervention in terms of the model of economic development—part of the universalistic tone of the campaign is that whatever your economic model corruption will still be bad for it, and hence "wrong". However, the specific reforms that TI sponsors are not part of all models of development but are similar to specific reforms featured in the specific reform programmes promoted by specific institutions of global governance.

From the anti-corruption campaigner's perspective (e.g. that of TI), this is just a coincidence. The key would be motivation: transparency is a good remedy against corruption, and it is a good prescription for attracting, for example, foreign investment. The anti-corruption campaigner's motivation is the former and not the latter. The fact that other campaigners with other agendas promote the same policies should not undermine the campaigner's altruistic motivation. Nevertheless, seen from neither the perspective of an anti-corruption campaigner nor of that of an advocate for liberalisation of trade, but rather from, say, the perspective of a civil servant of a developing country, then both these motivations seem to pile up, with the overall effect of delegitimising the exercise of discretion in policy making by states with poor economies. The point, then, is not that there is no theft of public resources in poor economies; of course there is. But the anti-corruption campaign suggests that there is an added value in thinking of these thefts (the "abuse of entrusted power for private gain" in TI's well-known definition) as more than criminally defined acts of embezzlement and rather as a systemic failure of governance that affects the very legitimacy and effectiveness of the corrupt state, with serious economic consequences. This systemic critique adds up to other narratives of systemic failures of governance (of failure of economic governance in the form of excessive regulation, for example, or of bizarre expenditure of public resources by dictators and their families), all of which end up sketching an overall landscape that features the state in poor economies as a suspect presence, in principle always working to rebut this presumption of failure.

However, this line of reasoning sounds somewhat rhetorical. While it seems plausible that this "piling up" phenomenon may occur, it seems hard

to measure. Similarly, while it is clear that the anti-corruption campaign coincides with important shifts in development policy as promoted by key institutions of global governance, particularly in their neoliberal mode, it is also hard to measure whether their commonality of agendas is just a mere coincidence of specific policies favoured by both neoliberals and anti-corruption campaigners, or whether there is deeper link of correlation or causality between the two. This is in stark contrast with the mammoth body of methodologically sophisticated literature that measures both corruption in the developing world and the correlation between corruption and economic under-performance. In a way, the argument that ties TI with its ideological and institutional milieu sounds subjective and perception-based, while the argument that explains the rationale for the anti-corruption campaign sounds objective and evidence based.

Some commentators have argued that this imbalance is a result of the existing consensus on the evils of corruption, as opposed to the lack of consensus on the evils of neocolonialism (Kennedy 1999, pp. 459–60).[5] However, there also seems to be a reason that is internal to the anti-corruption discourse. The anti-corruption campaign, and TI in particular, benefits from a general perception that corruption exists. At one level, the normative implication of this perception gives the basis for the general tone of the anti-corruption campaign that was discussed above—that is, its contention that corruption is wrong, everywhere. At another level, however, this perception is also an epistemic statement. While definitional problems abound, we seem to have here a general agreement on "facts" that is very rare in global governance: there is a general agreement that there are "facts" of corruption all around the world. TI is trying to react to these facts "on the ground"—regardless of the ideological implications that specific policies could entail. From these facts, the other two elements of the discourse (normativity and universality) seem to flow: if there is indeed corruption, and we can observe it as a fact, then the whole discussion of ideology seems irrelevant. The same standards should apply to both rich and poor economies: facts of corruption that we observe in poor countries would also be corrupt if they occurred in rich countries—and they should be normatively assessed with the same rigour.

This dependence on "facts" poses a particular challenge for the global anti-corruption campaign, and for TI in particular. On the one hand, it needs to produce a factual narrative of corruption that justifies its normative engagement. However, individual cases of fraud or embezzlement do not pose the structural challenge to economic development and public trust suggested by the anti-corruption campaign when it speaks of these problems—it is not a "banner wielding organization". In a telling division of labour, individual

cases are the concern of (mostly domestic) criminal jurisdictions, while the global anti-corruption campaign is concerned with wider patterns of corruption, the impacts of which can be felt structurally, by a country as a whole—or even globally. TI is focused on the macro dimension of corruption, while the micro cases are dealt with through criminal procedures. This means that the factual reality of corruption that needs to be constituted to this effect is not the same judicial "fact" that is required of the criminal prosecution. On the contrary, it needs to reflect a wider trend in the specific society—the factual reality of corruption needs to represent a social challenge, and not just the marginal deviance of criminal individuals.

This is the empty space that the CPI fills in this context. It constitutes the reality that is needed by the global anti-corruption campaign, at the level that requires it. As such, then, this role is perhaps more important than giving TI media visibility, or allowing it to name and shame particular countries into adopting reforms. The CPI constitutes a reality, based on a widely accepted intuition, and by doing so it sets the epistemic ground on which the normative and policy debates about corruption take place. To be sure, the reality that the CPI constitutes is just perception. However, criticising it as "inaccurate" or "false" would be to misunderstand the role of the indicator in this context. The reality that the anti-corruption campaign needs is not (and cannot be) composed of "real" facts of corruption, of the sort that are discussed in a criminal trial. On the contrary, the narrative of corruption as an obstacle to development requires precisely what the CPI is able to provide: a structural social challenge, which appears, essentially, in the same basic forms (grafts, bribes, etc.) all around the world, without exception. While individual corruption does exist, the kind of corruption that TI is keen on dealing with can only be observed through the prism of the CPI. In a way, the problem here did not trigger the tool to solve it, but rather the tool defined the problem it was supposed to solve.

Conclusion

This chapter has explored the use of indicators by civil society organisations in the context of the global anti-corruption movement. It has shown how using this technology opens new spaces for activism, yet poses difficult challenges that each organisation needs to assess in terms of its own organisational capacities. In the case of TI, given its background and resources, the benefits of doing activism through numbers seemed to outweigh the costs. This may not be the case for all civil society organisations, as the case of TI's national chapters seems to show. Beyond this specific balancing act, the CPI

plays a deeper epistemological role in defining the factual reality where TI's intervention occurs. This seems to be an important dimension of indicators: their power to frame issues is presumably a crucial resource for civil society organisations all around the world, regardless of their interest or particular agenda.

Notes

1. The 1996 Inter-American Convention Against Corruption, the 1997 OECD Convention on Combating Bribery of Foreign Public Officials in International Business Transactions, the 1999 (European) Criminal Law Convention on Corruption, the 1999 (European) Civil Law Convention on Corruption, the 2003 African Union Convention on Preventing and Combating Corruption, and, finally, the 2000 United Nations Convention Against Transnational Organised Crime.
2. On the transformation of the "Washington Consensus" from an economic policy for Latin America to a platform for development.
3. A crucial shift by the World Bank occurred in 1997, when the Board of Directors adopted the Bank's Corruption Strategy, which has been a key aspect of Bank policy ever since.
4. Calculations made with Google n-gram. http://books.google.com/ngrams.
5. For David Kennedy, for example, "the anti-corruption campaign benefits from the unwillingness of reasonable people to dispute plausible claims that corruption is taking place—the clear evil of the practice in general justifies at least some efforts to combat it. But the anti-anti-corruption campaign does not benefit from a similar generosity to the instinct that the campaign seems part of a broader ideological project of neo-colonialism. There is no countervailing general consensus about the evils of neo-colonialism, but rather quite the opposite. Charges of neo-colonialism seem vague and defensive efforts to change the subject, while charges of corruption seem straightforward efforts to get to the heart of things".

References

Alston, P. (2000). Beyond "Them" and "Us": Putting Treaty Body Reform into Perspective. In P. Alston & J. Crawford (Eds.), *The Future of UN Human Rights Treaty Monitoring* (pp. 501–525). Cambridge/New York: Cambridge University Press.

Andersson, S., & Heywood, P. M. (2009). The Politics of Perception: Use and Abuse of Transparency International's Approach to Measuring Corruption. *Political Studies, 57*(4), 746–767.

Clean, Not Laundered. (1993, May 8). *The Economist*. Retrieved from https://www.highbeam.com/doc/1G1-13811095.html

Cobham, A. (2013, July 22). Corrupting Perceptions. *Foreign Policy*. Retrieved from http://foreignpolicy.com/2013/07/22/corrupting-perceptions/

de Soto, H. (1989). *The Other Path: The Invisible Revolution in the Third World* (1st ed.). New York: Harper & Row.

Fjeldstad, O.-H., Isaksen, J., & Independent Evaluation Group – World Bank. (2008). *Anti-Corruption Reforms: Challenges, Effects and Limits of World Bank Support*. Washington, DC: World Bank.

Hollman, M. (1993, May 5). New Group Targets the Roots of Corruption. *Financial Times*, p. 4.

Kälin, W., & Künzli, J. (2009). *The Law of International Human Rights Protection*. Oxford/New York: Oxford University Press.

Kennedy, D. (1999). The International Anti-Corruption Campaign. *Connecticut Journal of International Law, 14*, 455–466.

Kingsbury, B., Merry, S. E., & Davis, K. E. (2012). Indicators as a Technology of Global Governance. *Law & Society Review, 46*(1), 71–104.

Lundberg, K. (2002). *High Road or Low? Transparency International and the Corruption Perception Index*. Cambridge, MA: Kennedy School of Government, Harvard University.

Mauro, P. (1995). Corruption and Growth. *Quarterly Journal of Economics, 110*(3), 681–712.

Mauro, P. (1996). *The Effects of Corruption on Growth, Investment, and Government Expenditure* (International Monetary Fund Working Paper No. 96/98). Washington, DC: International Monetary Fund.

McNickle, M. D. (1993, October 2). Group Set to Combat Third-World Graft. *The New York Times*. Retrieved from http://www.nytimes.com/1993/10/02/your-money/02iht-mrtrans.html?pagewanted=all

North, D. C. (1990). *Institutions, Institutional Change and Economic Performance*. Cambridge: Cambridge University Press.

North, D. C. (1991). Institutions. *Journal of Economic Perspectives, 5*(1), 97–112.

Norwegian Agency for Development Cooperation - NORAD. (2011). *Evaluation of Transparency International*. Oslo: NORAD.

Oman, C., & Arndt, C. (2006). *Uses and Abuses of Governance Indicators*. Paris: Development Centre of the OECD.

Open Contracting Partnership. (2016). Global Principles. Retrieved from http://www.open-contracting.org/global_principles

Pistor, K. (2012). Re-Construction of Private Indicators for Public Purposes. In B. Kingsbury, K. Davis, S. E. Merry, & A. Fisher (Eds.), *Governance by Indicators: Global Power through Quantification and Rankings* (pp. 165–179). Oxford: Oxford University Press.

Polzer, T. (2001). *Corruption: Deconstructing the World Bank Discourse* (LSE DESTIN Working Paper Series, 01-18, pp. 1–32).

Razzano, F. C., & Nelson, T. P. (2008). The Expanding Criminalization of Transnational Bribery: Global Prosecution Necessitates Global Compliance. *The International Lawyer, 42*(4), 1259–1286.

Sampford, C. J. G., Galtung, F., Shacklock, A., & Connors, C. (Eds.). (2006). *Measuring Corruption. Aldershot, England*. Burlington: Ashgate.

Santos, A., & Trubek, D. M. (2006). The Third Moment in Law and Development Theory and the Emergence of a New Critical Practice. In D. M. Trubek & A. Santos (Eds.), *The New Law and Economic Development: A critical Appraisal* (pp. 1–18). Cambridge: Cambridge University Press.

Satterthwaite, M. (2010). Indicators in Crisis: Rights-Based Humanitarian Indicators in Post-Earthquake Haiti. *New York University Journal of International Law and Politics, 43*, 865.

Satterthwaite, M. L., & Moses, P. S. (2012). Unintended Consequences: The Technology of Indicators in Post-Earthquake Haiti. *Journal of Haitian Studies, 18*(1), 14–49.

Shleifer, A., & Vishny, R. W. (1993). Corruption. *The Quarterly Journal of Economics, 108*(3), 599–617.

Transparency International. (2009). *CPI: Frequently Asked Questions* [Website]. Retrieved from http://www.transparency.org/files/content/tool/2009_CPI_FAQs_EN.pdf

Transparency International. (2016a). *Transparency International UK Charity Registration, No. 1112842*. London: Transparency International UK. Retrieved from http://apps.charitycommission.gov.uk

Transparency International. (2016b). *Current Operating Budget*. Berlin: Transparency International. Retrieved from http://www.transparency.org/whoweare/accountability/current_operating_budget/1/

Urueña, R. (2012a). Internally Displaced Population in Colombia: A Case Study on the Domestic Aspects of Indicators as Technologies of Global Governance. In B. Kingsbury, K. Davis, S. E. Merry, & A. Fisher (Eds.), *Governance by Indicators: Global Power through Quantification and Rankings* (pp. 249–280). Oxford: Oxford University Press.

Urueña, R. (2012b). *No Citizens Here: Global Subjects and Participation in International Law*. Leiden/Boston: Martinus Nijhoff Publishers.

Williamson, J. (2005). The Washington Consensus as Policy Prescription for Development. In T. Besley & N. R. Zagha (Eds.), *Development Challenges in the 1990s: Leading Policymakers Speak from Experience* (pp. 33–53). Washington, DC: World Bank Publications.

Wolfensohn, J. D. (2005). "People and Development". Address to the Board of Governors at the Annual Meetings of the World Bank and the International Monetary Fund. In A. Kircher (Ed.), *Voice for the World's Poor: Selected Speeches and Writings of World Bank President James D. Wolfensohn, 1995–2005* (p. 50). Washington, DC: World Bank.

17

Measuring Corruption in India: Work in Slow Progress

T.R. Raghunandan

Introduction

Over the past five years, the issue of corruption has received unprecedented attention in India. The revelation of a series of national- and state-level scams involving corruption in procurement[1] and licencing for the use of natural resources[2] from 2010 to 2012 resulted in civil society pressure seeking tangible action against the corrupt. These protests triggered widespread discussion in the media, Parliament and other public fora revolving around the causes of corruption, its different manifestations and how best to tackle it. The measurement of how corruption is perceived and experienced in India ought to have been central to this discussion; but it was sidelined by a general agreement that regardless of whether and how it was measured, corruption had become a chronic and all-pervasive disease infecting the public and private sectors in India. This chapter looks at the issue of whether, and if so how, corruption and its effects are measured in India. The focus of the chapter is on efforts that squarely focus on the issue of corruption. However, wherever relevant, attendant efforts to measure the quality of service delivery, which might obliquely look at corruption as an impediment to accessing services, are also examined.

T.R. Raghunandan (✉)
Avantika Foundation, Bengaluru, India

D.V. Malito et al. (eds.), *The Palgrave Handbook of Indicators in Global Governance*,
https://doi.org/10.1007/978-3-319-62707-6_17

Government Efforts

At the level of the national and state governments, there is no system for measuring corruption or its effects in India. At the national level, there are several reasons for this state of affairs. First, as India is a federal country, anti-corruption activity is a responsibility shared between the centre and the states and there is no hierarchical relationship between them in this regard. In addition, there is no clear responsibility bestowed on any institution, either at the central or state level, to measure corruption on a regular basis.

Second, at the central level, India does not yet have a single central ombudsman with the overarching responsibility for detecting, investigating and prosecuting corruption cases. There was nation-wide pressure between 2010 and 2012 for a powerful nation-level Ombudsman, the Lokpal, to be constituted. However, the nation-wide agitation dissipated after a group of agitators transformed themselves into a political party, the Aam Aadmi Party. Since then, a national law for the Constitution of the Lokpal, which bears significant differences from that presented by the agitators, has been approved by Parliament. The law, however, has yet to be implemented. Currently, the task of detecting and prosecuting corruption is also fragmented between several agencies, each examining different manifestations of corruption and aspects of its control, but none of which deal with the issue holistically (Box 17.1). None of these institutions have been entrusted with the responsibility for measuring corruption.

Box 17.1: Central Institutions Dealing with Corruption

The Central Vigilance Commission (CVC) looks at the detection and investigation of corruption in central government agencies. Reports by the Commission are sent to the government, which has the power to permit prosecution. The CVC is the closest institution to a national anti-corruption ombudsman in India.

The Central Bureau of Investigation (CBI) is a criminal investigation agency, which investigates several crimes, including corruption, either *suo moto*, or if they are entrusted to it.

The Department of Administrative Reforms and Public Grievances is a department of the Ministry of Personnel, Public Grievances and Pensions which is entrusted with the task of promoting and catalysing process change, which can improve efficiency and transparency and therefore reduce corruption.

The Directorate of Revenue Intelligence in the Ministry of Finance investigates, amongst other things, money laundering, which is a crime. Money that is "laundered" according to the definition of the Prevention of Money Laundering Act (which may or may not include money received as bribes) is liable to confiscation and forfeiture.

Source: Own compilation.

Third, another reason for the lack of an official government-driven system mandating a regular measurement of corruption is that India's overall framework for fighting corruption is quite weak. India's 1989 Prevention of Corruption Act only recognises 3 of the 12 types of corruption listed in the UN Convention against Corruption (UNCAC) as instances of criminally liable corruption under Indian law.[3] India was also one of the last countries to ratify the UNCAC and has yet to complete the first peer review following its ratification, which will examine India's efforts to align with the obligations resulting from its ratification of the UNCAC.

A similar situation exists at the state level, although the structure of the anti-corruption apparatus may vary from state to state. Unlike the centre, most states have an overarching anti-corruption Ombudsman referred to as the "Lokayukta". However, the Lokayukta is not vested with powers to undertake *suo moto* investigations or prosecute persons investigated. Whilst there are institutions that parallel the central apparatus, such as the State Anti-Corruption Bureaux,[4] the Corps of Detectives and the Departments of Administrative reforms, none of these are entrusted with or have developed a system for the measurement of corruption.

The idea of providing rights-based access to various services has been gathering momentum as an important step to strengthen the accountability of service providers. The first steps to rights-based access can be traced to the recognition of certain rights through the judicial expansion of the right to life enshrined in the constitution.[5] Beginning in the 1980s, the highest national courts have progressively and widely interpreted the constitutional provisions declaring that the right to life is much more than a mere right to exist; that citizens have enforceable rights for the state to provide them with a good quality of life. Another important step in the evolution of rights-based access to services is the emergence of the Right to Information Act, legislated a decade back.[6] The "National Campaign for People's Right to Information" (NCPRI), a campaign initiated in 1996, advocated the creation of the Right to Information Act in India. The bill came into force on October 12, 2005 and is viewed as one of the turning points in the history of Indian Democracy. The next step was the enactment of the National Rural Employment Guarantee Act (NREGA) in 2005, a law that guaranteed 100 days of assured labour to those who applied for it in rural areas. The enactment of the Right to Education Act and the food security bill are steps in furtherance of this trend. States have promulgated various rights to services laws in recent years. This legislative trend to enshrine access to services in law has now extended to cover a host of services. Recently, the NCPRI has also advocated a "Whistleblower's Protection Bill" and a "Grievance Redressal Bill".

Since 2006, the right to services has been legislated in 15 states and the central government has also come up with a national draft bill. These laws have a similar pattern, although details may vary from state to state. Essentially, they clearly articulate citizen entitlements and give a clear-cut assurance of the time frames within which services shall be delivered by designated service providers. Specifically, these Acts entitle citizens to demand and obtain relief in the case that services specified under the Act are not delivered within the stipulated time frame or are denied. While these laws do not explicitly mandate regular measurement of the quality of services, or of the presence, absence or extent of corruption in their delivery, they do build pressure for such measurement in future.

Civil Society and Private-Sector Efforts to Measure Corruption

It has therefore fallen upon civil society and other private-sector actors to take steps to measure several parameters that relate to corruption. Studies conducted by the India Chapter of Transparency International (TI) have been the most well known in this regard. Three such reports focussed on petty corruption have been prepared so far, in 2002, 2005 and 2008. The last two of these studies have improved upon previous efforts in terms of both coverage and methodology. They aim to go beyond perceptions to examine citizen experiences of corruption. The 2008 report presented its data in two parts. The first examined perceptions and experiences of corruption with respect to services and covered the delivery of public distribution, hospital services, school education, electricity, water supply, the national rural employment guarantee, land records and registration, forests, housing, banking and police services.[7] Part 2 of the report covered the same ground and more, but from a state perspective and contained state-wise reports on corruption.[8] On this basis the report also undertook a ranking of the states, from the least corrupt to the most corrupt. TI India also produces reports that are specific to a sector, for example, a study on corruption in setting municipal rates and one on the trucking sector in India (Box 17.2).

Another often-quoted example is that of the Public Affairs Centre (PAC) of Bangalore, which has developed a citizen's report card as a powerful toolkit for users of public services to provide public agencies with systematic feedback. The citizens' report card enables the collection of feedback from users on the quality and adequacy of public services and enables civil society organisations or local governments to engage in a dialogue with service

Box 17.2: Salient Features of Studies Conducted by Transparency International

Report on Corruption in Trucking Operations in India, conducted by Marketing and Development Research Associates (MDRA), TI (MDRA 2007):

- This study utilised qualitative and quantitative research methods to study corruption in trucking operations in India. 1222 truck drivers and operators were interviewed in 12 trucking centres in India.
- The study concluded that trucks pay between Rs. 211 and 266 in bribes per day. This amounts to Rs. 79,920 paid in bribes by a commercial truck during a year. India has 36 lakh truck operators which gives us an annual figure of Rs. 222,000 m spent on bribes in trucking operations.
- The majority of the bribes are paid at tollgates, check posts and state borders.

The India Corruption Study 2005 by Transparency International conducted a service-wise study of 11 Public Services, one of which was Municipal services (Centre for Media Studies 2005):

- The focus of the study was bribes paid by normal citizens to obtain services in municipal departments.
- The study concluded that bribes amounting to a value of Rs. 5500 m were paid in petty corruption.
- The average amount of bribes paid to the municipality was Rs. 2209 per household.
- More than two-thirds (69%) of the respondents who claimed to have paid a bribe had paid money to municipal officials directly, while 29% had paid money to agents/touts to get their work done.

Source: Own compilation.

providers to improve the delivery of public services. In 2014, PAC undertook a study on the Public Distribution System in Karnataka (Sreedharan et al. 2014) studying the effectiveness of its monitoring mechanisms by adopting its Citizen Card Methodology. PAC has also become a consultancy agency undertaking perception studies of public services on behalf of government departments. In 2016, PAC undertook an intensive study the outcome of which was the "Public Affairs Index" (PAI 2016). This index is to measure the quality of governance across all states in India, allowing state-wise comparison of the following parameters, namely, essential infrastructure, support to human development, social protection, women and children, crime, law and order, delivery of justice, environment, transparency and accountability, fiscal management and economic freedom.

Civil society's most successful initiative on the actual reduction of corruption is the Mazdoor Kisan Shakti Sangathan's (MKSS) efforts at widespread social audits of the NREGA. Starting in Rajasthan with the public questioning of officials and elected representatives on government spending, this effort was

galvanised by the enactment of the NREGA. Andhra Pradesh has pioneered a system where social audits are arranged and organised by a hybrid institution in which both the government and civil society participates. Drawing strength from the mandate contained in the Act that citizens' assemblies (Gram Sabhas) shall conduct social audits of activities they undertake, the state-level Social Audit Society facilitates a widespread performance of social audits by means of a large team of motivated volunteers and local activists. This has led to large-scale detection of misappropriated funds, possibly motivated by corruption, and forced their return into the government's coffers.[9] In a sense, detection of such irregularities reveals the extent of leakage, which again obliquely brings one closer to measuring corruption and its impacts.

A similar effort that has had a wide impact is the Annual Status of Education Report (ASER) survey undertaken by Pratham, an NGO that works on improving levels of education in India's schools. The annually conducted survey studies the extent of enrolment of children in school and the learning levels they attain by carrying out a simple test at the household level. What makes the approach credible is its rigour, consistency, wide coverage—more than 700,000 children are surveyed in all of India's 650 districts—and its citizen-friendly approach.

The "Planning, Allocations and Expenditures, Institutions: Studies in Accountability" (PAISA) project run by the Accountability Initiative is a pioneering Public Expenditure Tracking Survey in India. PAISA focusses on tracking fund flows to elementary schools and has a strategic partnership with the ASER so as to gain from the latter's capabilities and experience. Its surveys have enabled it to develop simple practical tools to collect data on fund flows, expenditures and implementation processes and develop and run training programmes that build capacities in citizens and the government to undertake tracking exercises and use the data to monitor service delivery. PAISA adopts a three-pronged approach: it produces annual reports based on its work in ten districts across the country covering 15,000 schools, which look at what happens to nationally funded programmes for universal education when these are translated into action at the sub-district and school levels. These reports provide "a quick snapshot of fund flows, implementation processes and an analysis of bottlenecks at the ground level" (Accountability Initiative 2016).

The second effort is to build capacity in the design and implementation of similar fund-tracking efforts. This has coalesced into a year-long course that covers the foundations of public finance, administration, basic statistics, communication and report writing. The third prong of the PAISA effort is to disseminate data and promote the use of PAISA data in local-level planning and decision-making. The project has been in operation since 2009 and is now

well recognised in the field of education as a reliable indicator of fund flow to schools and school administration institutions.

An interesting effort by civil society to measure corruption has been through the website ipaidabribe.com (IPAB 2016), which currently aims to bring the "market price" of corruption[10] to people's knowledge. The idea is that the site seeks and obtains details of bribes paid by citizens for a range of services and this would act as a market price tracker of corruption. If differences in the values of bribes demanded for the same service are tracked, it would throw up sufficient data to show whether the bribe being demanded in any one city or region was excessive. Therefore, one could discover if one was paying the right size of bribe based on the "market price" for that particular service in the corruption market. This approach has a sense of black humour and drew immediate attention to the site when it was started, encouraging citizens to visit the site and report their experiences on corruption. However, a more low-profile but useful objective of the site was to use citizens' reports on the nature, number, pattern, types, locations and frequency of actual corrupt acts and values of bribes as a knowledge bank that will contribute to a reduction in bribe payments.[11] Since achieving that objective was dependent upon the extent to which the key stakeholders — primarily government agencies concerned with the delivery of corruption-prone services — would respond positively, citizen reporting of bribe-related experiences dropped after the initial interest.[12] Changing the infographics, showing the movement of the market price of bribes on the site, did not slow or reverse this trend.

To counter the trend of falling interest and slowing down of the reporting of citizen's corruption experiences on the site, several measures were taken. A "Bribe Hotline" was established to make it easier for citizens to report corruption experiences. The hotline is a service where users can connect with experts and seek guidance on issues such as how to avoid paying a bribe, how to gain knowledge of the exact process or if need be call up a legal expert for guidance.[13] A partnership was forged with a law firm to help answer the legal questions. Following the introduction of the Bribe Hotline, there has been a dramatic increase in the number of reports recorded, to a total of 78,595 reports from 1071 cities by April 2016, involving a bribe value of 28.72 billion Indian Rupees.[14] This includes 39,478 reports received on the Hotline.

Nevertheless, challenges remain. From the standpoint of whether the "market price of corruption" constitutes a "measure" of corruption, the critical weakness of the site is that the approach of unrestricted crowd-sourcing of reports does not constitute a valid or unquestionable random sampling methodology. As a consequence, several defects creep into the measurement of the market price followed by the site. For instance, as it is a Bangalore-based

initiative operated by a well-known civil society organisation that operates from Bangalore, the majority of the reports are from that city and Karnataka, thus tilting the "corruption meter" adversely towards Bangalore and burdening the city with an undue disadvantage.

Second, the website does not indulge in data validation or verification and functions on the assumption that the reports posted on the site are true. As a result, there are instances where outliers and rogue reports continue to skew the analysis presented on the website.[15] It is therefore vital that crowd-sourced reports should be vetted on a regular basis to ensure that outliers and mischievous reports do not falsify the data.

The managers of the website agree that it would be a scientific mistake to use I Paid A Bribe (IPAB) reports as a correct reference for determining the market price of corruption and assert that the focus of ipaidabribe.com is not so much determining the economic market price of corruption, which is at best an inaccurate assumption by any standard, but being a place to communicate to and make the general public aware of the various modus operandi of the corrupt. They feel that the site serves as a warning by providing anecdotal reference for people to understand a grand unified picture of corruption experiences. They suggest that the "market price" could, however, be reconsidered as the cost burden of corruption on the individual by looking at the mean distribution of bribes across departments, which only serves as the least inaccurate measure of any form of cost valuation of corruption.[16]

In spite of the fact that its unstructured sampling methodologies reduce the "bribe meter" to nothing more than an attention-arresting curiosity, ipaidabribe.com still remains a unique and path-breaking effort with great potential. Its first outstanding feature is longevity. Similar sites started with the aim of tackling corruption when the anti-corruption effort was at its peak are either outdated or have lost their focus. Its second key feature is the persistence with which the site has responded to troughs of interest by citizens with several innovations and online campaigns to encourage them to continue to report their experiences on the site. Its third remarkable feature is the ease of replicability. With the proactive efforts of its organisers, the idea of ipaidabribe.com has been picked up and replicated in several other countries like Pakistan, Sri Lanka and Kenya.[17] Retention of the name and the broad approach of the site have resulted in brand recognition in the international arena of bribe fighting.

Ipaidabribe.com benefits a great deal from its relatively unstructured approach to inviting people to report their experiences with bribery. An increasing number of reports reflect an active citizenry looking for a platform to voice their distress. Moreover, there are no restrictions on the details that

an individual can recount on the site. Several of the incidents of corruption reported reveal a rich array of facts, process nuances and emotions of the actors that surround the bribe-giving act. When compiled and analysed, these provide valuable input for process reforms and psychological insights into the behavioural patterns of victims and perpetrators of corruption which it might not be possible to capture in a structured survey. However, for this potential to be fully realised there is a need of a competent research team which is able to study the data, draw inferences from it and provide inputs for process reforms and change management. That effort has happened with good results, but only intermittently (Box 17.3).

The current approach of the site is a three-pronged one: it redresses grievances by communicating with the bureaucracy and forming a conduit between IPAB reporters and the government; it provides the government with possible solutions; and it suggests process reform opportunities through research and inquiry and engaging in educational outreach with people in the age group 18–25 (currently through college and university outreach) in order to spread the awareness of corruption.[18]

As part of its effort to analyse the ease (or more appropriately, the difficulty) of conducting business in India, the consultant firm Ernst & Young conducts surveys on corruption, too. A recent report of their "Bribery and Corruption: Ground Reality in India" (2013) analyses the need for an anti-bribery and

Box 17.3: The "Janamahithi" Reports

When a sufficient number of citizen reports relating to a particular sector or transaction have been available for analysis, easy-to-read reports containing insightful analyses of departments and their transactions have been prepared showing the patterns of corruption and suggesting process changes that must be brought about to reduce the opportunities for corruption. These reports, termed "Janamahithi" (Jana = people, Mahithi = information) reports contain information on how citizens can avoid paying bribes by taking certain precautions, such as doing their homework and paperwork in advance and adopting certain styles of behaviour, and also process reform recommendations for the government. Two Janamahithi reports were prepared and presented to the Karnataka State government, one on corruption in the transport sector and the other on land registration. Some of the suggestions made in these reports were accepted by the government and appropriate process changes were made.

With respect to property sale transaction registration, the government adopted the concept of "anytime-anywhere" registration of properties. This resulted in a partial de-monopolisation of transaction registration offices, as customers could go to any office to have their property-related transactions registered.

Source: Own compilation.

corruption compliance programme for India along the lines of the US Foreign Corrupt Practices Act (FCPA) and the UK Bribery Act.

Benefits and Flaws of Civil Society and Private-Sector Efforts to Measure Corruption

This chapter will discuss important considerations concerning the production and use of corruption indicators in India. First, there is a demand for corruption indicators within Indian civil society. Civil society debate in India has been especially vibrant over the last seven years on the question of corruption. On the one hand, there have been large mass movements with much public participation, particularly from the urban middle class, which has otherwise been a passive bystander to debates on good governance in India. Barring a few exceptions, the public has largely lamented the generally falling standards in public life but there has not been enough debate on the process reform route to reducing corruption. A more data and research-driven approach to reducing corruption is required, so that public anger can be channelled into purposive action to reduce corruption, process by process.

The engagement of civil society actors in the development of instruments to measure corruption depends on the nature of their involvement in anti-corruption efforts. As detailed in the examples given above, some civil society agencies deal with corruption squarely. In individual cases, civil society's adoption of strategies such as social audits has yielded benefits in terms of corruption identified and corrected, and bribes and misappropriated money returned. However, the general tendency has been to approach corruption obliquely, as something that needs to be looked at within the larger context of accountability or transparency, or improving the efficiency of service delivery. Having said that, corruption indicators, even in their elementary form in India, have given civil society points of reference that support arguments for reform. Thus, if a certain service is identified as the "most corrupt" service, it helps to focus the attention of civil society, and through its pressure on policymakers, to take steps towards reform. Similarly, state-wise rankings have also put the spotlight on poorly performing states. However, since the measurement of corruption in India is limited and fragmented, these rankings have had a limited impact on decision-making and discourse-framing, both in the government and in civil society. Of course, there are exceptions, such as the healthy and frequent institutionalised interaction between civil society players and the government in the implementation of the NREGA.[19]

The downside of the Indian approach is that as existing corruption indicators are thin on the ground they provide little encouragement to local actors to engage in fighting corruption. The TI reports have been too infrequent to be considered an eagerly awaited national event to then give impetus to the debate. Individual studies do create a splash, but we are far from a situation where a report really represents a fork in the road, and where subsequent actions herald a new era. In such circumstances, world-level measures of corruption only provide the impetus for a general debate. As there is rarely enough comparable data generated from within the country, there is hardly any debate around the possible question of whether global measures of corruption conflict with the local-level understanding of corruption in the Indian context and the effectiveness of local anti-corruption initiatives.

As there are few robust corruption indicators in India, nothing more than speculation is possible on the likely impact of corruption on political, societal and cultural trends. In individual cases, successful efforts at detecting and reversing corruption have had impacts on political executives, such as in the NREGA. However, the impact of most reports on bribe-giving behaviour is probably poor. People read reports, sigh with despondence and go back to their bribe-giving ways. Even as positive changes might be happening, no ways have been developed to measure the extent and impact of these positive changes.

The absence of regular detailed national measurements of the incidence of corruption does not detract from interest in global-scale indices for democracy, legitimacy, delegation of authority and participation, particularly in view of India's rankings for these measures and the comparison of India with the state of progress in other countries. While the details of the composition of these indices and of how measurements are made might not be easily understood in their entirety by citizens, they are valuable in enhancing the quality and thrust of political debate and in prompting discussions in the media. The latter is particularly useful in India, which has a surfeit of TV channels that are often engaged in vibrant debate on these issues.

Since the development of corruption indicators and indices is in the formative stage in India, it would be premature to assess the potential benefits and/or pitfalls of their character to the fullest. The benefits of a "one size fits all" approach are that it enables a modicum of comparison between similarly placed entities. For instance, the state-wise ranking undertaken by TI India is easy to understand, provokes debate and is potentially influential—a chief minister of a state that is listed as highly corrupt will have to answer uncomfortable questions. However, when it comes to selecting the most effective interventions with respect to any particular service, suggestions need to be

made. In India, there are several variations in the manner of delivery of the same service across states. Baselines for service delivery not only vary between states but between regions within states too.

Indicators only enable a nuanced understanding of the problem of corruption. Some of the descriptive interviews undertaken through public surveys, citizens' report cards or crowdsourcing provide some indication of corruption-prone steps in service delivery workflows. These can enhance understanding of where reforms, such as automation of processes, will have the greatest effect.

It remains to be seen whether the globalisation of moral codes through corruption indicators has a significant influence in India, crossing over systemic, political and cultural differences. Moral codes vary at the margin; at their core they do not. However, the marginal variations do matter. For instance, gift-giving is a bribe in some cultures but not in others. Given these cultural nuances, corruption indicators will need to be anchored in a few global "must have" moral codes while recognising country- or culture-specific understanding of what corruption might or might not be.

Conclusion and Outlook

While much remains to be done in anti-corruption efforts in India, the ratification of the UNCAC is putting pressure on the government to take proactive steps to strengthen laws and policies in line with the Convention. These will include the strengthening of the law to enlarge the definition of corruption and reforming the institutional architecture at the central, state and local levels to combat corruption. Some of the steps taken to increase the accountability of state service delivery agents to people, such as the Right to Information Act, the Right to Services Acts and the various service guarantees such as the National Rural Employment Guarantee Act, the Food Security Act and the Right to Education Act also provide the impetus to civil society to undertake more comprehensive efforts to measure corruption, either across all services in a general way, or more specifically to burrow deeply into the delivery of a particular service.

However, the fact remains that there are no robust and credible approaches that focus on the measurement of corruption. Whatever is done is fragmented among several civil society organisations and private-sector consultants, often narrowly focussed on individual sectors or services. Wider, cross-cutting surveys such as the one undertaken by Transparency International are few and

not conducted at regular intervals. This is in marked contrast to the issue-based efforts at surveying the quality of service delivery aimed at propelling citizens to act and to put pressure on the government to improve its systems. Both the social audits spearheaded by MKSS and the ASER survey have captured the imagination of people and caused them to change their attitudes and demand better quality and more honestly delivered public services. India's experience also shows that when such pressures reach a critical mass the initial lassitude and hostility on the part of the government fade away. Internal champions begin to use the results of these surveys to catalyse reforms.

Given that India's legal frameworks for tackling corruption are in a state of transition, it would be timely to translate the relatively higher energy in the evaluation of service delivery efforts into a more pointed effort to track corruption. One way to do this is to promote thinking amongst the experienced campaigners in this regard — the Prathams and the MKSSs — to pioneer corruption measurement surveys. Efforts such as ipaidabribe.com could play an important role in the dissemination of information online. They could also leverage their online presence better if they were to refocus on their capability for quality research. However, that would require modification in their metrics of success, which will need to move on from website performance metrics to the extent to which they are influential in persuading the government to simplify, automate and "corruption-proof" government service-delivery processes. However, the downside of suggesting an entirely civil society-driven approach would be that it takes time to be recognised and taken seriously by the government. The MKSS's approach of street-level agitation and information-awareness campaigns took nearly two decades to bring about the change that they wanted to see; the enactment of the RTI legislation. The ASER approach, although more academic and distanced from public agitation, shows that for such a survey to gain credibility and acceptance by the government it takes a decade or more.

In such circumstances, the fundamental challenge is to develop sufficient numbers of survey methodologies to measure corruption; securing greater acceptance and credibility for them would be a second-order focus. The best way to create a credible and rigorous measure of corruption may be through creating a formal architecture for the conduct of a regular national survey (or a group of interrelated surveys) that focuses entirely on the issue of corruption. Would it be possible to build a coalition of government and civil society actors to conduct such surveys on a regular basis, or is this an unattainable utopian ideal? An approach that might work is to mandate the Lokpal, when constituted, with the task of conducting regular corruption incidence surveys. Initially, the Lokpal could focus on a few key sectors that encompass centre,

state and local responsibilities for service delivery. As experience builds up, the survey could extend to a wider range of services. This would be best organised as a decentralised effort with civil society, government actors and anti-corruption consultancies being entrusted with specific sectors to study. A peer review mechanism would ensure that standards remain consistent. Similar surveys could be organised at the sub-national level by Lokayuktas in states.[20] Giving the lead to the Lokpal at the national level and to the Lokayuktas in states would give such surveys the credibility that they deserve while ensuring sufficient autonomy to civil society and expert consultants tasked with the survey.

In conclusion, this is an aspect of governance in which India has a large distance to traverse. Whatever might be the approach adopted, there is a need to build capacities to undertake reliable measurements of corruption and to record the changes that take place over time in its character and nature. Current civil society efforts in India to measure the quality of service delivery such as ASER show that the capability to undertake large-scale surveys runs deep. These can provide inspiration and valuable insights to design similar efforts at tracking corruption, including the development of measures of corruption indicators which are appropriate in the Indian context.

Notes

1. For example, alleged corruption in the procurement of equipment for the Commonwealth Games held in India in 2010.
2. Two scams stand out in this regard. The first was triggered by a report by the Comptroller and Auditor General of India pointing out serious irregularities in the issuing of licences to telecommunication companies for the use of the broadband spectrum for provision of 2G mobile telephony services. The second related to widespread irregularities in the licencing and permitting of iron and manganese ore mining along the border between two states, Karnataka and Andhra Pradesh.
3. These are (a) bribery of national public officials; (b) embezzlement, misappropriation and other diversion of property by a public official; and (c) laundering of proceeds of crime. Bribery of foreign public officials and officials of public international organisations, abuse of functions, illicit enrichment, concealment, trading of influence and obstruction of justice; and bribery and embezzlement of property in the private sector (all listed in the UNCAC-2005) do not come within the definition of corruption in India. Moreover, companies in India are not liable for the corrupt acts of their employees. An amendment to strengthen the Prevention of Corruption Act and expand the

definition of "corruption" to include these acts was introduced in Parliament during the term of the previous government (2009–2014), but has not been passed yet.

4. Which may be merged with the Lokayukta or may exist as a separate entity.
5. Article 21 of the Indian Constitution states "Protection of life and personal liberty. No person shall be deprived of his life or personal liberty except according to procedure established by law."
6. RTI Acts quickly became the norm in many states in India before a national law was passed in 2005.
7. The following aspects of each service were investigated: interaction with the service concerned, purpose of interaction, difficulties faced during interaction, perception about the service concerned, measures taken to improve service, experience of corruption, reason for paying bribes, route of bribe payment, estimation of bribe amount, where do states stand, the service providers' perspective and suggestions to improve the service concerned.
8. Each state-wise chapter covered the following areas of investigation: general perceptions about corruption, interaction with public services, perceptions about services, grievance redressal mechanism, incidence of corruption, relative positioning of services, the service providers' perspective and the Right to Information & Citizens' Charter.
9. In 2006–07, 500 field assistants and 10 technical assistants were dismissed, 3 Mandal (Block—an administrative level above the village) officials were suspended and inquiries initiated against at least 6 other Mandal-level officials. Since the social audit process began, Rs. 6 million of embezzled funds have been returned (Aiyar and Samji 2009).
10. The present author was the Programme Coordinator for the site and in that capacity established it and ran it for nearly two years in 2010 and 2011.
11. ipaidabribe.com's original aim was to use and analyse the crowd-sourced reports by citizens to (a) heighten citizen awareness about the nature and diffusion of bribe-related exchanges and promote a purposive public debate to pressurise public officials to reduce and eventually eliminate corruption; (b) help citizens to recognise, avoid and tackle bribe-paying situations; and (c) identify and analyse the workflows within corruption-prone public services, resulting in suggestions on systemic reform directed at entrenching simpler and more transparent processes, more consistent standards of law enforcement, and better vigilance and regulation. In pursuance of the last objective, the site produced "Janamahithi" (peoples' information reports), which identified trends in bribery, the most corrupt offices and bribe-prone workflows and suggested process reform changes to reduce corruption. These were presented to the government with good effect. Another effort aimed at citizens was the preparation of the ten Commandments for Successfully Avoiding Corruption. These were drawn from reports by citizens who successfully resisted the payment of bribes.

12. On December 12, 2011, the site had had 16,725 experiences of citizens in a year and four months of online presence. In the next year and nine months, the number of bribe experiences only increased to 21,229 (4504 reports added, a 26% increase). As of today, (November, 10, 2014), the total number of reports stands at 31,701. However, the data reveal discrepancies: there are 20,284 reports of experiences of bribes paid, 2,573 where individuals did not pay bribes, but the number of cases where someone met an honest official is 884, bringing the total to 23,705 and not 31,701.
13. This commenced as a call by telephone going out to an external call centre, but now it is an online tool on the website answering questions regarding processes and legal recourses to counter corruption.
14. Details as of April 7, 2016. This includes 35,280 reports of bribes paid, 2,872 reports of bribe demands successfully resisted and 965 instances of honest officials who did not demand bribes.
15. For example, a bribe report reported in New Delhi (Is anybody doing anything?? 2014) was merely a rant about the nature of the website. However, the bribe amount added was Rs. 1,410,065,408, thereby leaving the city of Delhi with an absurdly high per-individual average bribe paid.
16. Email correspondence and response to a questionnaire sent by the author to the managers of the website in April 2016.
17. At the latest count, 30 countries have started similar sites based on the same brand.
18. Email response to a questionnaire from the managers of the site, April 2016.
19. In the NREGA, civil society actors are deeply involved in decision- and policy-making through various advisory bodies and task forces, which not only go into matters of high policy regarding the act of corruption but also into designing processes, such as software programmes.
20. Laws constituting Lokayuktas also task these bodies with addressing the question of maladministration. However, most of them function as grievance redressal and investigation authorities and do not pay attention to analytical studies of sectors with a view to bringing out sectoral white papers on corruption.

References

Accountability Initiative. (2016, December 16). *PAISA: Planning, Allocations and Expenditures, Institutions: Studies in Accountability*. New Dehli: Accountability Initiative. Retrieved from: http://demo.accountabilityindia.in/paisa-planning-allocations-and-expenditures-institutions-studies-accountability

Aiyar, Y., & Samji, S. (2009). *Transparency and Accountability in NREGA, a Case Study of Andhra Pradesh; Engaging Accountability* (Working Paper Series,

Accountability Initiative). New Dehli: Accountability Initiative. Retrieved from: http://www.accountabilityindia.in/paisa-planning-allocations-and-expenditures-institutions-studies-accountability

Centre for Media Studies. (2005). *India Corruption Study to Improve Governance*. New Delhi: Transparency International India. Retrieved from: http://www.transparencyindia.org/resource/survey_study/India%20Corruption%20Study%202005.pdf

Ernst & Young (EY) & Federation of Indian Chambers of Commerce and Industry (FICCI). (2013). *Bribery and Corruption: Ground Reality in India. A Survey by EY's Fraud Investigation & Dispute Services Practice*. Kolkata: Ernst & Young LLP. Retrieved from: http://www.ey.com/Publication/vwLUAssets/Bribery_and_corruption:_ground_reality_in_India/%24FILE/EY-FIDS-Bribery-and-corruption-ground-reality-in-India.pdf

I Paid A Bribe (IPAB). (2014). *Is Anybody Doing Anything??* I Paid A Bribe. Retrieved from: http://www.ipaidabribe.com/reports/paid/anybody-doing-anything#gsc.tab=0)

I Paid A Bribe (IPAB). (2016). [Website]. Retrieved from: http://www.ipaidabribe.com/#gsc.tab=0

Marketing & Development Research Associates (MDRA). (2007). *Report on Corruption in Trucking Operations in India*. New Delhi: Transparency International India. Retrieved from: http://www.transparencyindia.org/resource/survey_study/Corruption%20in%20Trucking%20Operations%20in%20India%20new.pdf

Public Affair Centre (PAC). (2016). [Website]. Retrieved from: http://pacindia.org/

Public Affair Index (PAI). (2016, December 16). *Public Affair Index. Governance in the State of India*. Retrieved from: http://pai.pacindia.org/

Sreedharan, S., Venugopala, R. A., Prabhakar, K., Srikant, P., & Harish, P. (2014). *Public Distribution System in Karnataka: A Study of the Effectiveness of Its Monitoring Mechanisms*. Bangalore: Public Affairs Centre.

18

Ranking Countries for Good Governance Using Public Opinion Surveys

Maksym Ivanyna and Anwar Shah

Introduction

In the last two decades there has been a proliferation of composite worldwide governance indicators purporting to measure various aspects of governance quality (see Arndt 2008 for the history and politics of governance ratings). The growth of these indicators has been spurred by generous support from the development assistance community, especially multilateral development finance agencies, international investors, and the infinite appetite of the media and the academic community for governance assessments and country rankings. Governance indicators are now being used as tools for conducting development dialogue, allocating external assistance and influencing foreign direct investment. For example, the World Bank's International Development Association (IDA) allocation—a window of subsidised lending to the developing world—and the US Agency for International Development's Millennium Challenge Account use various governance indicators as criteria for allocating external assistance. The post-2015 development agenda for Sustainable Development Goals gives even more prominence to progress in governance.

M. Ivanyna (✉)
Joint Vienna Institute, Vienna, Austria

A. Shah
Brookings Institution, Washington, DC, USA

D.V. Malito et al. (eds.), *The Palgrave Handbook of Indicators in Global Governance*,
https://doi.org/10.1007/978-3-319-62707-6_18

In view of the influential nature of governance indicators and their potential to do harm if the judgements they embody are biased or erroneous, it is imperative that they capture critical dimensions of the quality of governance and that all countries are evaluated using uniform and reasonably objective assessment criteria, which are consistent across countries and over time. The leading concerns in such measurement should be evaluation by citizens of the governance environments and outcomes in their countries, which should be supplemented by objective indicators and expert-based case study evaluations.

Do the existing indicators meet this test? There is an ample literature criticising the most widely used indicators, in particular the World Bank's Worldwide Governance Indicators (WGIs), for their lack of a conceptual framework on governance and time and cross-country inconsistency or opaqueness (see Arndt 2008; Arndt and Oman 2006; Kurtz and Schrank 2007; Iqbal and Shah 2006, 2008; Langbein and Knack, 2008; Schrank and Kurtz 2008; Thomas 2006; Ivanyna and Shah 2011). However, one of the most important limitations common to all the available composite indexes of governance is that they do not capture how citizens perceive the governance environment and outcomes in their own countries. Most indexes are either solely based on external expert evaluations or when these are mixed with citizen evaluations they usually give the latter much less weight. For example, in the WGIs 2013 only 8 of the 150 representative data points are based on public opinion surveys. If we simply average these eight data points and compare the resulting country ranking with that of the WGI itself (an average of six aggregate indicators that the WGI project reports), the mean absolute percentile difference would be 24, that is, with 215 countries covered, an average country would be expected to change its ranking up or down by about 50 positions. This shows that citizens in many countries seem to hold very different opinion about their governments to those of international experts.

In this chapter, we lay out our proposal for a uniform and consistent global framework for measuring governance based on citizen evaluations. A comprehensive data source that is consistent across countries and over time and that is suitable for implementing this framework in practice has yet to be developed. For demonstration purposes and to gain a better understanding of the practical details, we put our framework into action by using the World Values Survey, which combines decent geographical coverage with an acceptable number of governance-related questions. We compute citizen-centric governance indicators for 100 countries over 6 waves of the survey from 1980 to 2014—more than once for many countries.

Public opinion contains useful and unique information for measuring governance, but it is harder to interpret than objective indicators or expert assessments. It is often subject to systemic biases, stemming, for example, from indoctrination by the mass media or government oppression, which invalidate comparison across countries and over time. The good news is that in carefully constructed public opinion surveys these biases can be corrected, at least partly. In this chapter we suggest several ways to do this. Our practical exercise suggests, however, that even after the adjustment the results of citizen-centric indicators can be quite different from those of conventional measures of governance for some countries.

The remainder of the chapter is organised as follows. Section "Conceptualising and Measuring Governance in a Comparative Context" presents a summary of the literature on the concept of governance and its measurement and clarifies the debate on whether or not comparative measurement should focus on governance processes, as many worldwide governance indicators do, or governance quality outcomes as, for example, Huther and Shah (1996) do, or on both, as proposed by Ivanyna and Shah (2011) and this chapter. Section "Towards a Citizen-centric Framework for Assessing Country Governance Quality" specifies a citizen-centric conceptual framework for measuring governance quality. Section "Citizen-centric Governance: Empirical Framework" presents an empirical framework, data sources and aggregation techniques. Section "Citizen-centric Governance: Preliminary Results" presents the preliminary results. In section "Removing Systematic Biases from Public Opinion" we discuss the robustness of the results, in particular a correction for systemic biases in public opinion. A concluding section outlines an agenda for future research. The general aim of this chapter is to provide an overview of this line of argumentation. More details can be found in Ivanyna and Shah (2011, 2015).

Conceptualising and Measuring Governance in a Comparative Context[1]

Governance is a fuzzy yet fashionable buzzword and its use in the literature has exploded in recent years. According to the American Heritage, Random House and Merriam Webster dictionaries, governance is equated with government and is defined as the "exercise of authority and control" or "a method or system of government and management" or "the act, process or power of governing". Huther and Shah (1996, 1998) define governance as "a multi-faceted concept encompassing all aspects of the exercise of author-

ity through formal and informal institutions in the management of the resource endowment of a state. The quality of governance is thus determined by the impact of this exercise of power on the quality of life enjoyed by its citizens". Kaufmann et al. (1999, p. 1) define governance as "the traditions and institutions by which authority in a country is exercised". The main difference between these definitions of governance is whether they focus on governance processes (governance as input), or governance outcomes (governance as output).

Both process-based and outcome-based definitions of governance are useful, but none is perfect. Definitions with a singular focus on processes/institutions do not lend themselves to easy or fair comparability across countries and sometimes do not do so even within one country without deeper analytical studies being conducted. There can be little disagreement that the same processes and institutions can lead to divergent governance outcomes, just as dissimilar processes could yield similar outcomes in two different countries. For example, anti-corruption agencies in countries with fair governance help curtail corruption but those in countries with poor governance prove either ineffective or, worse, a tool for corrupt practices and victimisation (Shah 2007). During recent decades, we have also seen that dominant-single-party political systems in China, Malaysia and Singapore have shown dramatic results in improving governance outcomes whereas pluralistic party systems have also shown positive results in other countries such as Brazil and India. Comparisons of processes and institutions out of their context are often ideologically driven and value laden and cannot be acceptable as unbiased professional (scientific) judgements. In any case, deeper analytical work through comparative studies rather than aggregate process indicators is required.

Governance outcomes also assume commonly shared values, but it is relatively less problematic to agree on these than on "one size fits all" prescriptions on processes. At the same time, outcome-based measures of governance are likely to be less actionable, that is, provide less guidance on where to improve. Moreover, the border between processes and outcomes is sometimes blurred. For example, should the availability of elections be viewed as a process or an outcome?

To have meaningful governance comparisons across countries and over time, there is a need for concepts which are somewhat invariant over time and place and are focused on citizen evaluations that capture both the quality of governance processes and their outcomes. The following section clarifies the approach taken by this chapter.

Towards a Citizen-centric Framework for Assessing Country Governance Quality

From a neo-institutional perspective, various orders of government (agents) are created to serve, preserve, protect and promote public interest based on the values and expectations of the citizens of a state (principals). The underlying assumption is that there is a widely shared notion of the public interest. In return, governments are given coercive powers to carry out their mandates. To this end, we follow Ivanyna and Shah (2011) and define governance as "an exercise of authority and control to preserve and protect public interest and enhance the quality of life enjoyed by citizens". Note that this definition encompasses both the governance environment (quality of institutions and processes) and governance outcomes.

A stylised view of the public interest can be characterised by four dimensions of governance outcomes:

- *Responsive Governance*. The fundamental task of governing is to promote and pursue collective interests while respecting formal (rule of law) and informal norms. This is done by government creating an enabling environment for the right things to be done—that is, it promotes and delivers services consistent with citizen preferences. Furthermore, the government only carries out the tasks that it is authorised to do, that is, it follows the compact authorised by citizens at large.
- *Fair (equitable) Governance*. For peace, order and good government, the government mediates conflicting interests, is focused on consensus building and inclusiveness and ensures a sense of participation by all and protection of the poor, minorities and disadvantaged members of society.
- *Responsible Governance*. The government does it right, that is, governmental authority is carried out following due process with integrity (absence of corruption), with fiscal prudence, with concern for providing the best value for money and with a view to earning the people's trust.
- *Accountable Governance*. Citizens can hold the government to account for all its actions. This requires government to "let sunshine in" on its operations and to work to strengthen the voice of the principals and their exit options. It also means that government truly respects the role of countervailing formal and informal institutions of accountability in governance.

Given the focus on governance outcomes, Table 18.1 presents some preliminary ideas for discussion on how to operationalise these concepts in individual country assessments.

The above simple framework captures most aspects of governance outcomes, especially those relevant for development policy dialogue, and can serve as a useful starting point for a consensual framework to be developed. Once such a framework is developed, one needs to only focus on a few key indicators that represent citizen evaluations and that can be measured with some degree of confidence in most countries of the world and can be defended for their transparency and reasonable degree of comparability and objectivity (see Andrews and Shah 2005; Shah and Shah 2006).

Implementation of the above framework requires a worldwide survey with a uniform questionnaire homing in on the four dimensions of governance identified above across countries. Such a survey has yet to be developed. In the

Table 18.1 Governance outcomes and relevant considerations

Governance outcomes	Relevant considerations
Responsive governance	• Public services consistent with citizen preferences; • Direct, possibly interactive, democracy; • Safety of life, liberty and property; • Peace, order, rule of law; • Freedom of choice and expression; • Improvements in economic and social outcomes; • Improvements in the quantity, quality and access to public services; • Improvements in the quality of life.
Fair governance	• Fulfilment of citizens' values and expectations in relation to participation, social justice and due process; • Access of the poor, minorities and disadvantaged groups to basic public services; • Non-discriminatory laws and enforcement; • Egalitarian income distribution; • Equal opportunity for all.
Responsible governance	• Open, transparent and prudent economic, fiscal and financial management; • Working better and costing less; • Ensuring the integrity of its operations; • Earning trust; • Managing risks; • Competitive service delivery; • Focus on results.
Accountable governance	• Justice-able rights and due process; • Access to justice, information; • Judicial integrity and independence; • Effective legislature and civil society oversight; • Recall of officials and rollbacks of programmes possible; • Effective limits to government intervention; • Effective restraints on special interest capture.

Source: Shah 2008, p. 25

following section, we take a pragmatic approach based upon available survey data to develop rough indexes of governance quality in order to test the framework in practice.

Citizen-centric Governance: Empirical Framework

The procedure of the citizen-centric governance assessment consists of two main steps. The first step is to choose a data source—the individual responses from an intercountry public opinion survey. The survey must, at least to a certain degree, contain questions, which characterise governance outcomes as in Table 18.1. The second step is to aggregate the responses into a governance index for each country in the sample.

Step 1: Data Selection

Data that are reliable, comprehensive and consistent over time and space are essential for the qualitative estimation of citizen-centric governance indicators (CGIs). With the additional requirement of being publicly accessible and, preferably, free of charge, such a data source hardly exists at present. There is, however, a database of governance-related questions included in different surveys around the world (Governance Surveys Database, published by the World Bank). In principle, each of these questions (taken separately from different polls) could be included in the estimation if the data are available. However, as experiments with the construction of surveys suggest (see Bertrand and Mullainathan 2001 for examples), even a small difference in the formulation of a question (assigned to the same sub-criterion) or in a sequence of questions in a survey may lead to significant discrepancies in the responses for the same country and the same sub-criterion. Therefore, it is essential to only use one data source which covers a sufficient number of countries and provides governance-related questions which are consistent over time.

Table 18.2 summarises the main potential data sources that could be used to assess CGIs. None of the sources is perfect, in fact quite the contrary. In the absence of such a perfect survey, of all the possible sources we consider that the World Values Survey (WVS) provides an acceptable compromise between consistency and coverage to show an initial picture of CGIs (see WVS 2015). Although it publishes quite outdated information (with a time lag of two to three years after the actual survey was made) with only a few questions relevant for our purposes (since the survey is mainly about cultural values, not

Table 18.2 Existing sources of data and their main features

		Geographical coverage				Data access		
Name	Code	Number	Region	Years	Av. frequency	Free	Time lag	Relevance of questions
World Values Survey	WVS	100	Worldwide	1981–2014	3–6 years	Yes	2–3 years	Average
Afrobarometer	AFR	30	Sub-Saharan Africa	2001–2015	3 years	Yes	1–2 years	High
Asiabarometer	ASB	25	East Asia	2003–2008	2 years	Yes	1–2 years	High
Business Environment and Enterprise Performance Survey	BEEPS	29	Central and Eastern Europe	1999–2014	3 years	Yes	1–2 years	Low
Transparency International Global Corruption Barometer	GCB	107	Worldwide	2004–2013	1 year	Yes	Less than 1 year	Very low
Latinobarometro	LBO	18	Latin America	2004–2013	1 year	No	1 year	High
Eurobarometer	EUB	30	Europe	1973–2015	Half-year	Yes	Less than 1 year	Very high
Gallup World Poll	GWP	160	Worldwide	2007–2014	1 year	No	n.a.	n.a.
GWP—datapoints from WBI (2015)	GWP WGI	119	Worldwide	2007	1 year	Yes	Immediately	Low

Note: *Number*—the total number of countries which participated in the survey; *Time Lag*—the time period between making the survey and posting data; *Relevance of questions*—correspondence of questions in the questionnaire to the sub-criteria for governance in Table 18.1 on the scale very low/low-average/high/very high (authors' opinion)

Source: own compilation

governance), it provides quite comprehensive geographical coverage (100 countries, with all the major economies included) combined with acceptable time coverage (six waves, from 1981 to 2014).

Table 18.3 presents the questions in the WVS that could be (and are) used to measure governance. However, for several of the sub-criteria in

Table 18.3 Governance outcomes: weights and questions assigned

Code	Governance criteria	Questions assigned	Weight
A	Responsive governance		
1	Safety of life, order, rule of law	How much confidence do you have in the police?	0.035
3	Improvements in economic and social outcomes	How satisfied are you with the financial situation of your household?	0.15
4	Improvements in the quality of life: general	All things considered, how satisfied are you with your life as a whole these days?	0.1
5	Improvements in the quality of life: health	All in all, how would you describe your state of health today?	0.07
6	Peace	How much confidence do you have in the armed forces?	0.035
7	Improvements in the quality of life: happiness	Taking all things together would you say you are [happy, unhappy]?	0.1
B	Fair governance		
1	Social justice, respect for human rights	How much respect is there for individual human rights nowadays in the country?	0.08
2	Government represents the whole country	How proud are you to be your nationality?	0.035
3	Government represents the whole country	Would you fight for your country?	0.035
C	Responsible governance		
1A	Earning trust: executive branch	How much confidence do you have in the government?	0.075
1B	Earning trust: legislative branch	How much confidence do you have in the parliament?	0.075
2	Earning trust: general	How much confidence do you have in the civil service?	0.07
D	Accountable governance		
1A	Access to information, independent mass media—press	How much confidence do you have in the press?	0.035
1B	Access to information, independent mass media—television	How much confidence do you have in the television?	0.035
2	Judicial integrity and independence	How much confidence do you have in the courts?	0.07

Note: The data source for all questions is World Values Survey (WVS). The questions are common to all six waves of the survey

Table 18.1 no survey questions are available. Many governance-related questions are not asked in all waves or in all countries. However, at least ten of the sub-criteria for governance summarised in Table 18.1 are covered by questions with sufficient representation over time and across countries.

An interesting feature of the resulting dataset is that the observations can be sorted by the gender, income and education of the respondent, as well as by the sub-national administrative unit of his/her residence and other characteristics. On the basis of these disaggregate variables the corresponding "adjusted" CGIs can be constructed.

Step 2: Aggregation

The assumption underlying our empirical investigation is that the quality of governance in a given country directly affects the governance outcomes analysed in a certain survey question. Thus, the better the answers by the survey respondents—citizens of the country—to each question are, the higher the quality of governance in the country is. At the same time, the respondents' answers are random variables which are subject to personal errors:

$$s_{ijk} = g_i + \varepsilon_{ijk} \Rightarrow g_i = s_{ijk} - \varepsilon_{ijk},$$

where $i = 1 \ldots M$ is the country index, $j = 1 \ldots N_i$ is the respondent index (obviously, the total number of respondents changes from country to country) and $k = 1 \ldots K$ is the index of particular questions in a survey (thus of a particular governance outcome). s_{ijk} is the answer to question k by respondent j in country i. Each response is normalised to the range 0–1, with 0 being the worst answer, and 1 being the best answer. g_i is the quality of governance in country i, which obviously does not depend on either the individual respondent or on a specific question. Finally, $\varepsilon_{ijk} \overset{i.d.}{\sim} N\left(\mu_{ik}, {\sigma_{ik}}^2\right)$ is the random error which is assumed to be independently normally distributed with mean μ_{ik} and variance σ^2_{ik}, and both may depend on the country and the specific question. If μ_{ik} is zero, then s_{ijk} provides an unbiased estimate of g_i. This is the assumption that we make in this and the next sections. If μ_{ik} is not zero, then a systemic bias in public opinion is present, and it has to be corrected.

Given our assumptions, the most efficient unbiased consistent estimator for governance in country i is just the sample mean of the weighted averages

of the citizens' responses. The estimator for the variance in governance is the adjusted sample variation:

$$\hat{g}_i = \frac{1}{N_i}\sum_{j=1}^{N_i}\sum_{k=1}^{K} w_k s_{ijk}, \hat{\mathrm{var}}(g_i)$$
$$= \sum_{k=1}^{K} {w_k}^2 \frac{1}{N_i - 1}\sum_{j=1}^{N_i}\left(s_{ijk} - \frac{1}{N_i}\sum_{j=1}^{N_i} s_{ijk}\right)^2,$$

where the weights w for each question are chosen to minimise the variance in the governance indicator. Roughly speaking, questions with smaller variance σ in the measurement error ε should receive greater weight. Since σ's are not observed, the eventual choice of weights is effectively arbitrary. We take a parsimonious and comprehensive approach and assign equal weight to each sub-criterion in Table 18.1 "covered", with the exception of a few questions which seem to be relatively more far-reaching in their assessment of governance (e.g. "satisfaction with life in general" is clearly more comprehensive than "satisfaction with health"). Such questions receive greater weight. All the weights add up to one.

It is up to the researcher to use more sophisticated data-mining approaches (e.g. principal component analysis or random projections) to choose weights, but one needs to be transparent and simple if the resulting indicators are to be understood by the general public. The procedure that we adopt here is maximally open and simple in order to allow for further research and analysis. Moreover, we also report and analyse the responses to each separate question, which allows conclusions to be drawn which are completely independent of the weights and aggregation procedure.

Citizen-centric Governance: Preliminary Results

Maps of the citizen-centric governance indicators for wave 5 (2005–2009) and wave 6 (2010–2014) of the survey are presented in Figs. 18.1 and 18.2. The surveys were made in 57 and 59 countries, respectively. The countries are grouped in four quartiles: the darker the colour, the higher the CGI.

Several observations emerge from Figs. 18.1 and 18.2. As expected, most developed countries (especially the Scandinavian countries, Switzerland, Canada and New Zealand) demonstrate stable and good performance. At the same time, a number of developing countries are among the top performers, which may appear surprising to a reader from an OECD country but can be justified as these governments have an excellent track record in improving

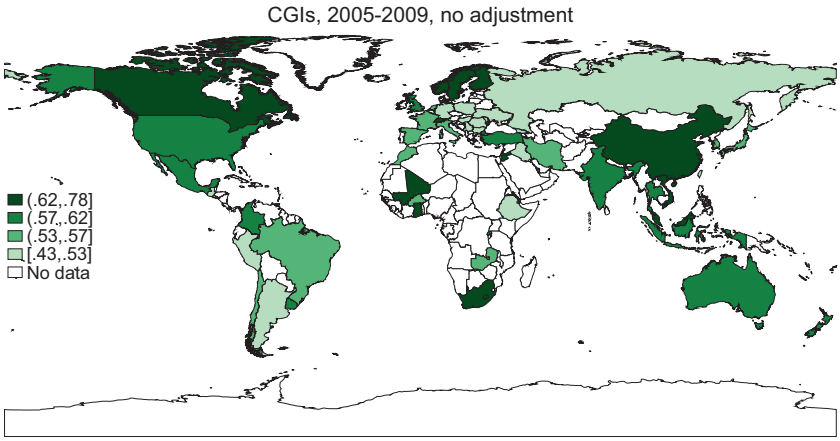

Fig. 18.1 Citizen-centric governance indicators 2005–2009. Source: Data source WVS, waves 5 and 6

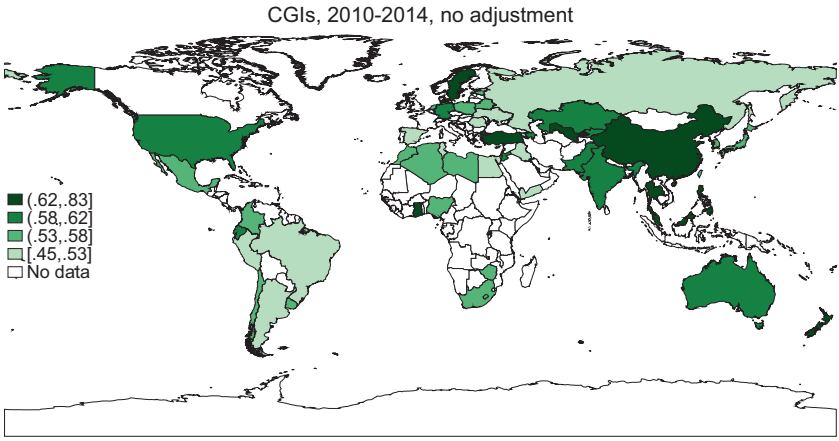

Fig. 18.2 Citizen-centric governance indicators 2010–2014. Source: Data source WVS, waves 5 and 6

economic and social outcomes for their residents and this is captured in their citizens' evaluations. East Asian governments (especially Vietnam and China) get particularly high rankings from their population. Ghana, South Africa and Turkey also perform highly according to local public opinion. On the other hand, countries in Central and Eastern Europe and Latin America consistently get the lowest scores.

Figure 18.3 shows that in general there is positive correlation between citizen-centric governance indicators and "conventional" measures of governance (e.g. WGI Control of Corruption index). The fit, however, is far from

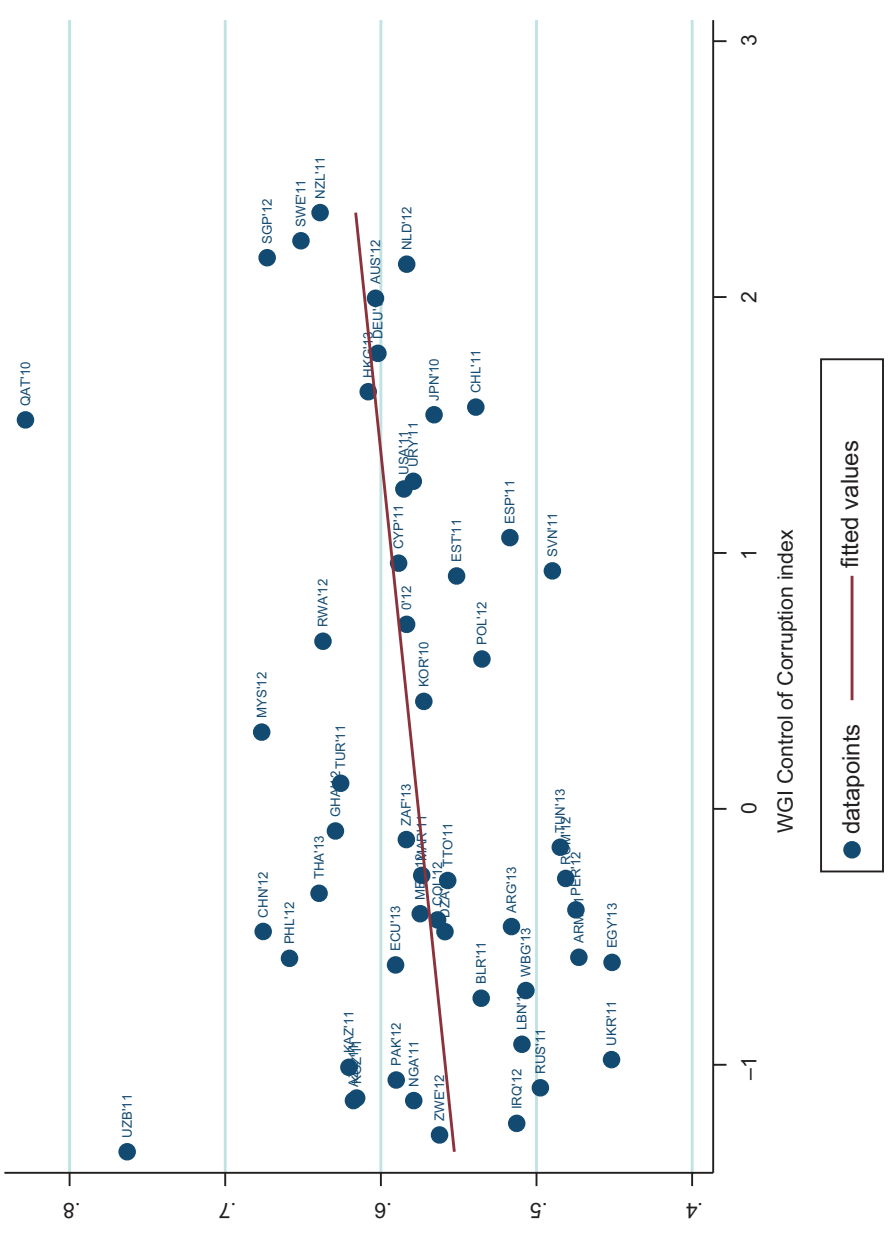

Fig. 18.3 Citizen-centric governance indicators vs. WGI Control of Corruption index, wave 6. Source: WGI Control of Corruption index (WBI 2015), CGI wave 6

perfect. Many countries which are ranked low by the WGI get high scores from the CGI. One reason for this is that public opinion contains information about governance institutions and outcomes which is not present in expert-opinion-based measures. Another potential reason could be that public opinion is subject to systematic biases. We explore this in section "Citizen-centric Governance: Preliminary Results".

CGIs: Extensions

Working with raw survey data provides ample opportunities for adjusting and calibrating governance measures to the researcher's needs. First, to avoid arbitrariness of weighting one can analyse CGIs question by question. Second, CGIs are perfectly appropriate for tracking governance over time, as the selected questions remained unchanged throughout the six waves of the survey. Third, CGIs can be computed over various sub-samples. For example, one can compute how governance is perceived among the poorest or the richest, among people with higher education or by sub-national units. The latter property is unique to public-opinion-based measures, and is a distinct advantage over expert-opinion-based measures. We demonstrate a couple of examples below.

Figure 18.4 shows the dynamics of CGIs and selected components between the fifth and the sixth wave of the WVS. The common sample over both waves consists of 36 countries, both developing and advanced. According to

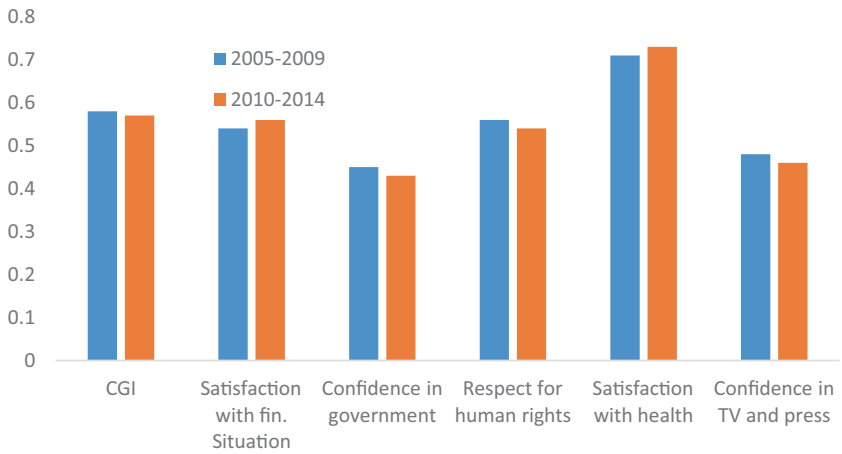

Fig. 18.4 Citizen-centric governance indicators: Time dynamics. Source: Authors' calculations

the figure, quality of governance does not seem to change much between 2005 and 2014. Between 2010 and 2014, people became relatively more satisfied with their health and financial situation—most probably the outcome of the 2008–2009 financial crisis. At the same time confidence in government and in the mass-media went down. Governance progress by country reveals an interesting pattern. The top ten worst performers include countries like Cyprus and Slovenia, which recently experienced major financial crises, and countries like Brazil and Ukraine, which eventually faced massive social protests.

Figure 18.5 demonstrates another application of CGIs—a comparison between how governance is perceived by poor (lowest 30th percentile) and rich citizens (highest 30th percentile). Poor people consistently gave their governments lower scores than the rich. At the same time, the gap between these scores, from being relatively small in the 1990s, began to widen in the 2000s: from 0.06 points at the beginning of 1990s to 0.15 points in 2010–2014. This worrying trend is consistent with the rising income inequality around the world, and the ever more frequent calls from civil society and international organisations to make economic growth more inclusive.

CGIs can also be computed at the sub-national level. For instance in Germany in 2006, the rich industrial lands[2] of Hessen, Nordrhein-Westfalen and Saarland together with the independent cities of Bremen, Hamburg and

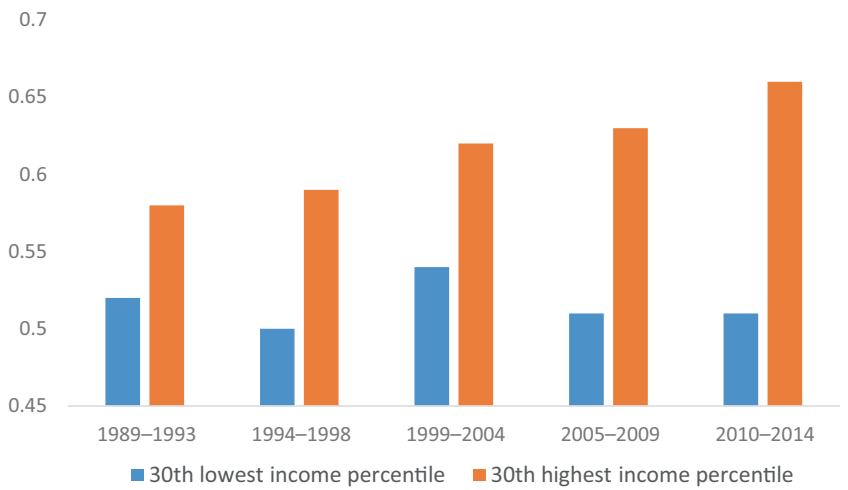

Fig. 18.5 Citizen-centric governance indicators by income groups: 30th lowest income percentile vs. 30th highest income percentile. Source: Authors' calculations

Berlin were the most satisfied with their governments. At the same time, the scores were much lower in the poorer eastern part of the country—only in Sachsen-Anhalt did citizens give their government more than 0.55. The surprising results are that the scores received by the governments of rich southern states—Baden-Würtemberg and Bayern—were quite moderate. The relative correspondence between the richness of a jurisdiction and its CGI score is also found for Italy. In 2006, all the regions except one in the country's rich north scored more than 0.55 while five out of the nine regions in the poorer south and centre had a result below 0.55 (see details in Ivanyna and Shah 2011).

These results for sub-national CGI represent, to our knowledge, the first attempt to assess governance at a less aggregate level than that of the country. Analysis of such results may prove helpful in empirical research on decentralisation and governance, decentralisation and welfare, differences between capital and non-capital regions, industrialised and rural regions, and so on.

Removing Systematic Biases from Public Opinion

As mentioned above, the relatively poor fit between CGIs and expert-opinion-based measures of governance may be caused by two factors. First, CGIs may contain information about governance which is simply not present in other measures. Second, public opinion may suffer from various kinds of systematic biases. Systematic biases distort the real picture of citizens' assessment of governance and therefore have to be removed. In this section we discuss possible ways to do this. Our conclusion is that biases are indeed present, but even after adjustment CGIs still contain new (compared to expert-opinion-based measures) information about governance.

Public opinion—especially about issues related to the government—might be influenced by at least three systemic biases, which have to be removed. First is the "intimidation" effect, where people are afraid to express their true negative opinion about their government because they think they could be punished for it. The second factor is the "indoctrination" effect, where the mass media in a country praise or criticise the government disproportionately so that it distorts public opinion. The third factor which may bias public opinion is the degree of citizen activism and the perceived role of government in a country. In particular, Norris (1999) argues that in the 1970s there was an emergence in developed countries of a class of so-called critical citizens—people, who were more and more critical and demanding towards their governments despite their obvious successes.

Assuming that "intimidation", "indoctrination" and "critical citizenship" affect respondents' answers as well as the actual quality of governance, it is not possible to estimate the magnitude of these effects since the governance is not observed. However, the problem can be resolved if we note, or rather assume, that for some questions (call them "objective") the effect of "intimidation", "indoctrination" or "critical citizenship" is likely to be very close to zero, and for some the effect is likely to be strong. For instance, when an individual is asked about satisfaction with her/his health, it is likely that she/he will not be intimidated to tell other than the truth. At the same time, questions like "Do you have confidence in your government?" (call them "subjective") are most probably subject to all the above-mentioned biases. Therefore, we can isolate the effect of biases on responses to individual questions by considering the difference between "subjective" and "objective" questions.

Our empirical model and estimation procedure to assess the magnitudes of "intimidation", "indoctrination" and "critical citizenship" effects are described in detail in Ivanyna and Shah (2015). Here, we only present the results.

To adjust CGIs for indoctrination we start by measuring individual exposure to the mass media—by looking at the frequency with which an individual exposes her or himself to TV, the press and internet. Specifically, we use the questions "How often do you watch TV?", "How often do you read newspapers?" and "How often do you use Internet?" from the WVS.[3] The more people watch TV, read newspapers or use the internet the more likely they are to be exposed to possible indoctrination or excessive criticism if they are present in the mass media. To check if they are indeed present, and to what degree, we run country-specific regressions for subjective-objective question differences in exposure to the mass media and other individual-specific variables. The larger the coefficient is, the more indoctrination there is; and a negative coefficient means that the mass media is excessively critical.

Our results show that both statistically significant indoctrination and excessive criticism are indeed present in many countries. In most of these, television seems to be the primary source of indoctrination. Everything else being equal, watching TV on a daily basis was likely to bias public opinion about governance by around 0.03 points on average in 2010–2014. The coefficient is as high as 0.1 in Tunisia. Reading news on the internet, on the contrary, seems to make people more critical of their governments, but the absolute value of the bias is smaller than in the case of TV. The average coefficient is −0.016 (vs. 0.03 for TV). Local newspapers are on average neutral, but there is much heterogeneity among countries. Our main conclusion is that even though citizens in many developing countries seem to be indoctri-

nated, mass media bias is also present in many developed countries—New Zealand, Japan, Sweden, Australia, the USA. The degree of indoctrination (the coefficient on mass media exposure) may be relatively lower in developed countries, but the total bias can be quite large because of higher exposure (e.g. 0.07 in New Zealand, or 0.06 in Japan). The average bias in 2010–2014 is 0.022, and it varies from −0.035 in Thailand to 0.091 in Tunisia.

Intimidation and "critical citizenship" biases are identified in cross-country regressions. As a proxy for the intimidation level, we use the average score of the country in the "Freedom in the World" ranking—an annual publication by Freedom House which assesses citizens' political and civil rights. As for "critical citizenship", we follow Pippa Norris (1999) in her definition of a "critical citizen" and define a country to be in the stage of "critical citizenship" if it had been classified as "free" by Freedom House for at least ten years before the survey was conducted (long period of stable democracy) and the current GDP per capita is more than 10,000 US dollars (wealthy population). Most OECD countries enter this group. To identify intimidation and "critical citizenship" biases we include both variables in a cross-country regression where the dependent variable is the subjective-objective question difference, averaged over all the respondents in a country.

We find that both the freedom of the county and its "critical citizenship" status are statistically significant in explaining the biases of subjective questions in the WVS surveys. The directions of the effects are as expected. One score higher in the Freedom House ranking (which means the country becomes less free, 1 being the best score, and 7 the worst) does make people more cautious in answering government-related questions in a public opinion survey, and consequently they over-praise their governments by 0.013 points. On the other hand, residents of countries which are in a stage of "critical citizenship" do have significantly less confidence in their governments then they should have. If they were not "critical", the residents of these countries would have given their governments a score 0.039 points higher. The total CGI adjustment for both biases varies from −0.09 for Uzbekistan to 0.026 for most OECD countries.

Figures 18.6 and 18.7 report the adjusted CGIs for 2010–2014 and a comparison with the WGI Control of Corruption index, respectively. The first thing to note is that the adjusted CGIs are now much closer to conventional measures of governance. Western European countries, the USA, New Zealand and Australia are now in the upper half of the ranking. Latin American countries improve their rankings somewhat. The relationship

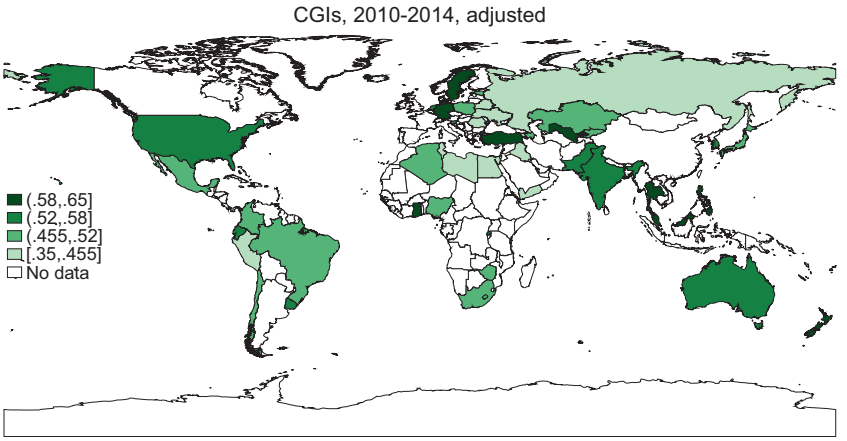

Fig. 18.6 Citizen-centric governance indicators, wave 6, adjusted for indoctrination, intimidation and critical citizenship. Source: Authors' calculations

between adjusted CGIs and WGI Control of Corruption is clearly positive and highly significant. Nevertheless, even after the adjustment CGIs still seem to contain information which is not captured by other governance indexes. The scores for East Asian countries, Uzbekistan and Ghana go down, but these countries still remain in the upper half of the country ranking. Apparently, there are other reasons for some governments to score so high in public opinion polls. In the case of East Asia, the main reason is probably the stable economic growth and development in the region in the last decade, as Wang (2005) argues for China. At the same time, poor economic performance, political conflicts and corruption in the 1990s (and for many countries until the present day) in Central and Eastern European countries keep the scores of the governments in these regions quite low (although Poland shows significant progress in the last wave).

Contributions and Limitations of the Empirical Approach

Measuring governance on the basis of public opinion surveys is important. In this chapter we have provided a conceptual framework for doing this consistently across countries and over time. We have also illustrated the empirical implementation of the framework using data from the World Values Surveys. The rankings we obtain, while positively correlated with expert-opinion-based

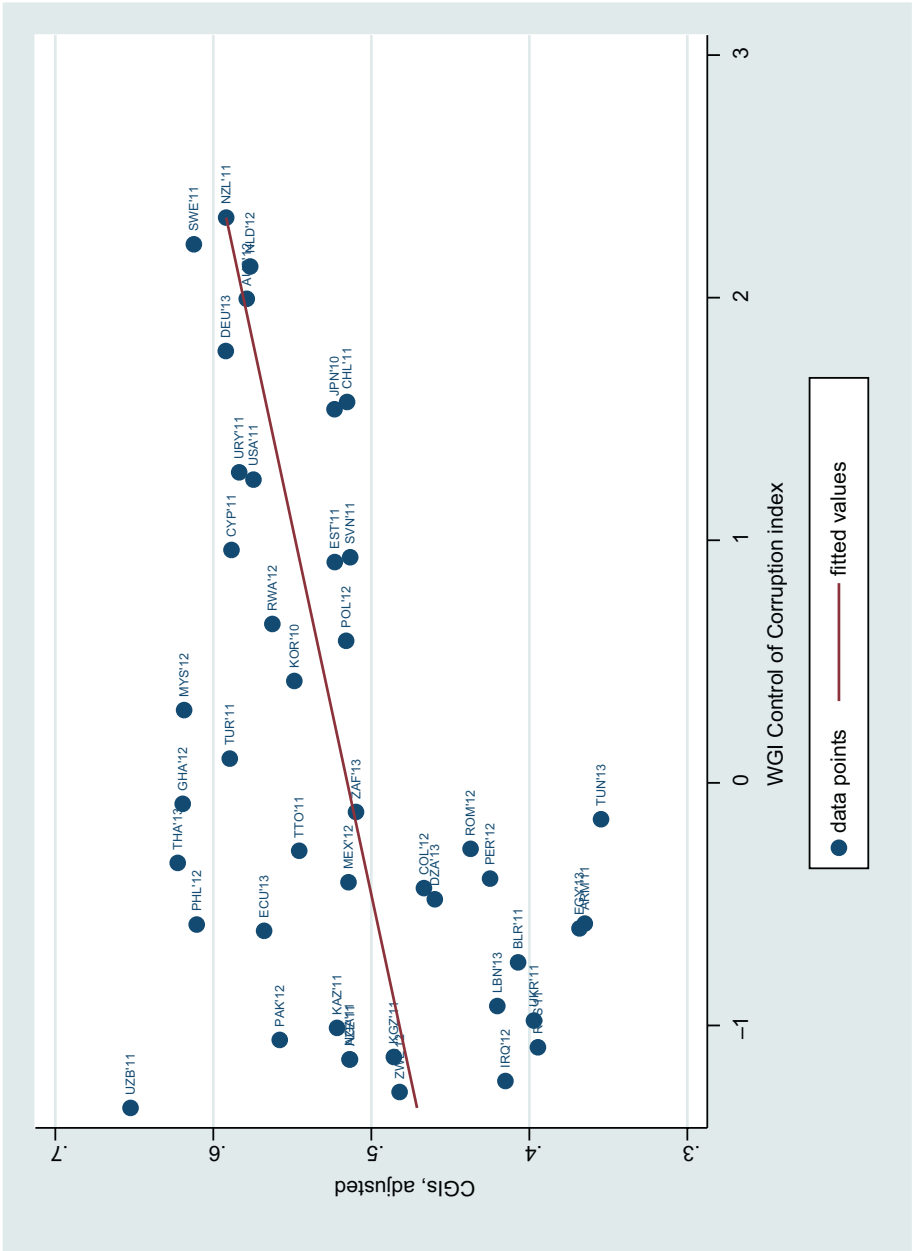

Fig. 18.7 Adjusted citizen-centric governance indicators vs. Worldwide Governance Indicators, wave 6. Source: WGI Control of Corruption index (WBI 2015), CGI wave 6

measures, differ greatly from these measures. The differences become smaller after we adjust CGIs for potential systemic biases of indoctrination, intimidation and critical citizenship, but still remain significant. This means that public-opinion-based governance measures carry information about governance which is not captured by other measures.

Even though we have suggested ways to adjust CGIs for potential biases that might be present in public opinion, surveys, in particular those conducted within the WVS project, are certainly subject to important limitations. For example, they are not conducted annually for all countries and the questionnaires may differ slightly from country to country, which may produce significant departures from objective estimation. It is also possible that, in spite of the claims to the contrary by the survey organisation, the survey may not be based on stratified random sampling for some countries due to practical difficulties. Public opinion surveys could also be subject to other systemic biases which we have not accounted for—for example, the general level of optimism in society, or a general cultural tendency to respond positively. Other methods should be developed to adjust for these biases (see our discussion in Ivanyna and Shah 2015).

Notwithstanding these limitations, the WVS dataset has important merits for a citizen-centric approach to measuring governance. The governance-related questions and answers are reported at the level of individual respondents, which gives researchers great flexibility in composing rankings. In particular, it is possible to compose rankings for groups of citizens with higher education, different genders, income, and so on. Most importantly and unlike the WGIs, the data that we have used in our estimation is freely accessible and can be easily used by other researchers to replicate or modify our estimation procedure.

Ideally, our theoretical framework should be implemented using a world poll with stratified random sampling employing a uniform questionnaire across countries and over time. The World Gallup Poll or a similar instrument might offer such an opportunity in the near future.

Notes

1. This section draws heavily upon Ivanyna and Shah (2011) and Iqbal and Shah (2008).
2. Länder in German—second-tier jurisdictions in the country.
3. The exact formulation differs from wave to wave.

References

Andrews, M., & Shah, A. (2005). Citizen-Centered Governance: A New Approach to Public Sector Reform. In A. Shah (Ed.), *Public Expenditure Analysis* (pp. 153–182). Washington, DC: World Bank Publications.

Arndt, C. (2008). The Politics of Governance Ratings. *International Public Management Journal, 11*(3), 275–297.

Arndt, C., & Oman, C. (2006). *Uses and Abuses of Governance Indicators*. Paris: Development Center Studies, OECD.

Bertrand, M., & Mullainathan, S. (2001). Do People Mean What They Say? Implications for Subjective Survey Data. *American Economic Review, 91*(2), 67–72.

Huther, J. & Shah, A. (1996). *A Simple Measure of Good Governance* (Unpublished Paper, Operations Evaluation Department). Washington, DC: World Bank.

Huther, J. & Shah, A. (1998). *Applying a Simple Measure of Good Governance to the Debate on Fiscal Decentralization* (Policy Research Working Paper 1894). Washington, DC: World Bank.

Iqbal, K., & Shah, A. (2006). *Critical Appraisal of Governance Perception Indicators and Their Uses*. Washington, DC: World Bank Institute. Retrieved from http://www.worldbank.org/wbi/publicfinance.

Iqbal, K., & Shah, A. (2008). *Truth in Advertisement: How Do Governance Indicators Stack Up?* Washington, DC: World Bank Institute. Retrieved from http://www.worldbank.org/wbi/publicfinance.

Ivanyna, M. & Shah, A. (2011). *Citizen-Centric Governance Indicators: Measuring and Monitoring Governance by Listening to the People* (CESifo Forum, 1/2011, pp. 59–71). Munich, CESifo GmbH.

Ivanyna, M. & Shah, A. (2015). *Towards Greater Objectivity in Governance Measurement: Second Generation Citizen-Centric Governance Indicators*. Mimeo. Retrieved from: https://msu.edu/~ivanynam/research/cgi2/Ivanyna_Shah_CGI2_2015.pdf

Kaufmann, D., Kraay, A. & Zoito-Lobaton, P. (1999). *Governance Matters* (World Bank PRD Working Paper, No. 2196). Washington, DC: World Bank.

Kurtz, M., & Schrank, A. (2007). Growth and Governance: Models, Measures and Mechanism. *The Journal of Politics, 69*(2), 538–554.

Langbein, L. & Knack, S. (2008). *The Worldwide Governance Indicators and Tautology* (Policy Research Working Paper, No. 4669). Washington, DC: World Bank.

Norris, P. (1999). *Critical Citizens: Global Support for Democratic Governance*. New York: Oxford University Press.

Schrank, A. & Kurtz, M. (2008). *Conceptualizing and Measuring Institutions: A View from Political Science* (Unpublished paper). Columbus: Department of Political Science, Ohio State University.

Shah, A. (2007). Tailoring the Fight Against Corruption to Country Circumstances. In A. Shah (Ed.), *Performance Accountability and Combating Corruption* (pp. 233–254). Washington, DC: World Bank Publications.

Shah, A. (2008). Demanding to Be Served: On Holding Government to Account for Service Delivery. In J. de Jong & G. Rizvi (Eds.), *The State of the Access*. Washington, DC: Brookings Institution Press.

Shah, A., & Shah, S. (2006). The New Vision of Local Governance and the Evolving Role of Local Governments. In A. Shah (Ed.), *Local Governance in Developing Countries* (pp. 1–46). Washington, DC: World Bank.

Thomas, M. (2006). *What Do the Worldwide Governance Indicators Measure?* (Unpublished Paper). Washington, DC: John Hopkins University.

Wang, Z. (2005). Before the Emergence of Critical Citizens: Economic Development and Political Trust in China. *International Review of Sociology, 15*(1), 155–171.

World Bank Institute. (2015). *Worldwide Governance Indicators*. Washington, DC: World Bank Institute. Retrieved from: http://info.worldbank.org/governance/wgi/#home.

World Values Survey. (2015). *1981–2014 Longitudinal Aggregate v.20150418*. Madrid: World Values Survey Association, JDSystems. Retrieved from: www.worldvaluessurvey.org

19

Analysing the Use of Sustainability Indicators

Stephen Morse

Introduction

This chapter focuses on one of the key tools in the sustainable development toolbox that can allow those trying to steer society towards a more sustainable route to know whether they have made the right choices, namely, the use of sustainability indicators (SIs) and indices. There are many ways of assessing progress in sustainable development, but SIs have become a popular approach (Rametsteiner et al. 2011; Dahl 2012; Turcu 2013; Pissourios 2013). SIs attempt to crystallise complex data sets into a framework of related indicators, or a single measure in the case of an index, that can be readily grasped by non-specialists and thus acted upon. Unlike other approaches, such as Life Cycle Analysis (Jeswani et al. 2010; Guinee et al. 2011), the use of SIs for this purpose seeks to bridge the gap between technical specialists involved in research and collecting data and those who make the decisions. Underpinning all of this is a sense of the "right" direction of travel or, put another way, the definition of sustainable development that should be adopted. While this is important, the chapter will not seek to delve into these bigger issues of definition and the gaps between rhetoric and action; other sources do this in a way that space does not permit here (see, e.g., Blowers et al. 2012; Leeuw et al. 2012).

Given that sustainable development is a multiverse with many dimensions, it should not be surprising that there are many SIs designed to cover aspects

S. Morse (✉)
University of Surrey, Surrey, UK

D.V. Malito et al. (eds.), *The Palgrave Handbook of Indicators in Global Governance*,
https://doi.org/10.1007/978-3-319-62707-6_19

of these dimensions (Bell and Morse 2008). There is no single global initiative or, indeed, set of standards for developing SIs, and in reality, they have been produced by many individuals and groups, including international agencies, political parties, governments, civil society groups, the media and private companies. Each of these groups may have a particular interest to promote, and they see SIs as an aid to help achieve this, but the interests of different groups may collide rather than be synergistic or even compatible. The result is a complex and diverse "ecosystem" of SIs, some of which overlap in terms of their focus and some share the same data; some evolve over time and some have even been discontinued. A single SI, such as the Human Development Index (HDI), published by the United Nations Development Programme (UNDP) since 1990, may undergo many changes in terms of its components and how they are assembled and calculated into a single value (Morse 2013, 2014). Indeed, the HDI published in 2012 has little resemblance to that published in 1990, even if the name and underlying ethos have remained the same. The SI multiverse is thus a complex space of SI creation, evolution and death, all mediated by interest groups trying to exert an influence on others. The dynamics within this "SI multiverse" are driven by humans, their wishes and their desires. Despite the quantitative nature of SIs and the maths used to create them, this is certainly not a space of pure objectivity. SIs are, of course, human constructs to help meet the perceived (by humans) needs of humans (Morse 2004).

One type of SI is indicators of fiscal sustainability (fSI). This is a rather narrow set of SIs designed to assess the sustainability of economies—typically those of nation states. An example is the ratio of debt to gross domestic product (GDP), usually expressed in percentage terms. The higher this value is, the lower the assumed "fiscal sustainability" of an economy, in effect suggesting that the economy is having to rely on borrowed money for recurrent payments and/or investment. The values for this fSI vary a great deal, and some examples for a number of European states are shown in Fig. 19.1; they range from 140% (Greece) to 8.5% (Estonia). The polarity in Fig. 19.1 implies that fiscal sustainability improves from the left to the right-hand side of the graph, but it should be noted that this is an interpretation founded on just one fSI.

In contrast to the extensive literature focussed on more technical aspects of SI development, there have been relatively few attempts to assess their value in terms of their influence on policy or indeed interventions in general (Bell and Morse 2011). Part of this is, no doubt, due to the fact that dissecting a presumed cause-effect link between an SI and a specific policy or intervention is challenging. The SI may be just one element amongst many influences that have an impact on the development of a policy, and separating out this

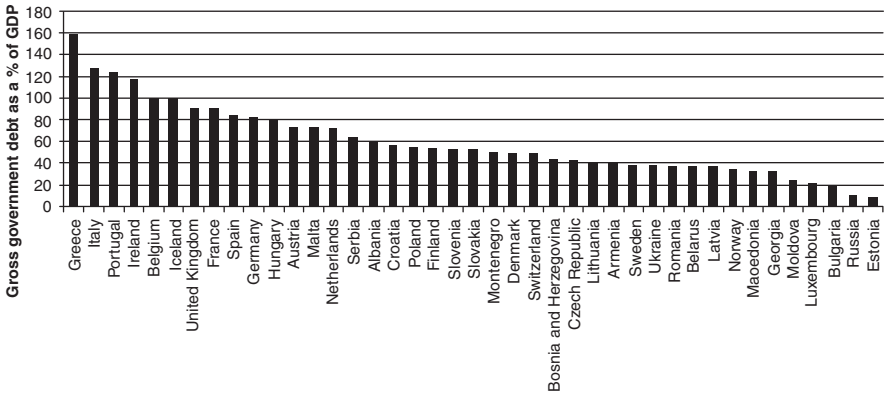

Fig. 19.1 Debt-to-GDP ratios for some European countries. Source: own compilation based on International Monetary Fund (2012) data

contribution from all the others may not be straightforward. Herzi (2004) suggests that there are various ways in which SIs can be used:

- *Instrumental*: indicators inform decisions that have impacts.
- *Conceptual*: they catalyse learning and understanding.
- *Tactical*: they are a substitute for action and deflect criticism.
- *Symbolic*: they provide ritualistic assurance.
- *Political*: they support a pre-determined position.

The first of these—instrumental use—is perhaps the one that most people think of with regard to SIs and is perhaps the one which SI developers and promoters tend to have in mind. Instrumental use implies that an SI is measured on a regular basis and its value has a direct influence on what managers do. For example, if an SI reaches a certain threshold, then the manager may switch to a different set of interventions. The second in the list—conceptual use—implies that an SI can still be "used" even if it does not play a direct role in policy or intervention. Once an individual becomes aware of an SI, it helps catalyse a wider understanding of the concept that the SI is attempting to measure. For example, an individual may come across the HDI and this may raise awareness of the concept of human development. The other three uses suggested by Herzi (2004) are perhaps less positive. For example, tactical use refers to situations where someone may selectively quote an SI, perhaps out of context, as a means of deflecting criticism or perhaps to imply that nothing needs to be done, that all is well.

A European Union (EU)-funded research project called "Policy Use and Influence of Indicators" (POINT), completed in 2011, remains one of the

few in-depth studies of SI use, and some of the findings of the project are summarised in a special edition of the journal "Ecological Indicators" published in December 2013. The project came up with many insights, but most notably that

> perceiving the role of indicators in purely 'instrumental' terms as an input to rational policy making is insufficient. Indicators are not necessarily influential just because they are available, technically plausible, required and used in some sense. (Frederiksen and Gudmundsson 2013, p. 2)

However, it is the instrumental use of SIs that is often regarded as their predominant role. The POINT project suggested that

> When instrumental use was in fact observed to a limited extent, it seems to be favored by the pre-existence of structured policy problems and frameworks, policy agendas with a certain weight and stability, and binding goals or objectives that are regularly monitored. (Frederiksen and Gudmundsson 2013, p. 2)

Thus, it would seem that an instrumental use of an SI requires the right environment for that to happen. Given that governments tend to prioritise economic growth over many other areas of policy, SIs linked to fiscal sustainability may be expected to be "valued" by them and play an instrumental role that should also be obvious to others. However, this is based on a somewhat sweeping assumption and no studies exist as yet which show that this is the case for fSIs, or indeed explore the factors that could encourage or discourage such a use. These are the questions that form the basis for the work described here.

Read All About It: Measuring the Use of SIs By the Press

One approach to exploring SI use would be to look at reporting by the media—a group that can be presumed to have some influence over the general public and, indeed, civil servants, politicians and so on (Holt and Barkemeyer 2012; Mekelberg 2012; Schmidt et al. 2013). This is admittedly something of a complex two-way relationship as the media is, in turn, influenced by these same groups (Barabas and Jerit 2009), but it seems reasonable to assume that if the media "uses" an SI in its reporting, then that is a measure of "success", even if a very limited one. It has previously been reported that the

press do make use of SIs—for a variety of reasons (Morse 2011a, b). While newspaper articles can be focussed on an SI, often following the release of a new report that presents values of the SI, typically in a "league table" format such as a listing of countries by the HDI, they may also use SIs to highlight wider issues. For example, an article may look at issues of child labour or civil strife in a country, and the reporter may use the HDI as an illustration of the development standing of that country. Similarly, an article may be about corruption and use the Corruption Perception Index (CPI) and/or the Bribe Payers Index (BPI) as an illustration of how this may vary between countries and which of them may be "hot spots". Thus, the type of use of SIs may vary a great deal.

One advantage of exploring use in the context of newspaper reporting is that commercial search engines exist which allow the number of articles that mention an SI at least once to be counted. This search facility is not available for other types of media (TV, radio etc.) or indeed for sources such as policy documents. Newspapers are available in many countries and in many languages. They are a relatively inexpensive type of medium, at least when compared to TV, radio and the internet. On the other hand, of course, they require a degree of literacy, which is a significant disadvantage when compared with TV and radio. Nonetheless, the relative ease of being able to carry out content analysis of newspaper articles is a major advantage.

In the research reported here, the content analysis that was undertaken was of a fairly basic type—essentially searches for the number of articles that mention an SI at least once in the text. There are a number of problems with this approach, of course, including the fact that it does not take into account the intensity of use in an article. Thus, articles which mention an SI just once, maybe in passing, are counted the same as those that have a more intensive use of an SI to illustrate a point being made. Second, it does not take into account the context in which the SI was used—a point noted above. Was the SI used as an ornamental "add-on", or was it linked more closely to the fabric of the article? Third, it is possible to have a high degree of reporting of an SI by just one newspaper. This is significant, of course, but the diversity of readership may perhaps be limited. It is known that consumers will typically stick with one or perhaps a few newspapers, and this can be influenced by a number of factors, including socio-economic grouping, so for an SI to have the widest possible exposure to the public, it should ideally be used by as wide a variety of newspapers as possible. This latter point can be accommodated by assessing the "diversity" of reporting of an SI, and this is described in the next section.

Selecting a sample of SIs in order to assess their reporting by the press and to make a comparison among SIs for fiscal sustainability is a challenging task.

The decision was taken to select as large a sample as possible (in this case 24; Table 19.1) using a set of defined criteria, although the latter are subjective, and thus open to some criticism. The SIs were taken from a survey undertaken by the UNDP (Bandura 2008), which lists a total of 178 "measures" designed to assess country performance in various aspects of development, but it was decided to focus only on measures that are discrete (single numbers) and only on those that had been updated over a number of years rather than being just "one-off" creations.

It should be noted that the Bandura (2008) report does not necessarily label all the measures in Table 19.1 as SIs, but here it will be assumed that those selected cover important aspects within the commonly accepted conceptual framework of sustainable development (Fig. 19.2) and hence will be referred to as SIs.

Figure 19.2 sets out the three main spheres of sustainable development, and the space where they overlap in the centre is usually taken as the point of

Table 19.1 Examples of sustainability indicators

Index	Acronym	Starting year	Number of years reported
Big Mac Index	BMI	1988	25
Bribe Payers Index	BPI	1999	14
Carbon Footprint	CF	2001	12
Climate Change Performance Index	CCPI	2006	7
Commitment to Development Index	CDI	2003	10
Corruption Perception Index	CPI	1996	17
Debt to GDP ratio	None	1986	28
Democracy Score	DS	2004	9
Ecological Footprint	EF	1993	20
Environmental Performance Index	EPI	1996	17
Environmental Sustainability Index	ESI	2000	13
Failed States Index	FSI	2005	8
Gender Empowerment Index	GEI	1995	18
Gender Development Index	GDI	1995	18
Genuine Progress Indicator	GPI	1994	19
Global Competitiveness Index	GCI	1996	17
Global Hunger Index	GHI	2007	6
Global Peace Index	GlPI	2007	6
Happiness Index	HI	1987	26
Happy Planet Index		2006	7
Human Development Index	HDI	1990	23
Human Poverty Index	HPI	1997	16
Living Planet Index	LPI	1998	15
Mothers Index	MI	2001	12
Press Freedom Index	PFI	2002	11

Source: Own compilation

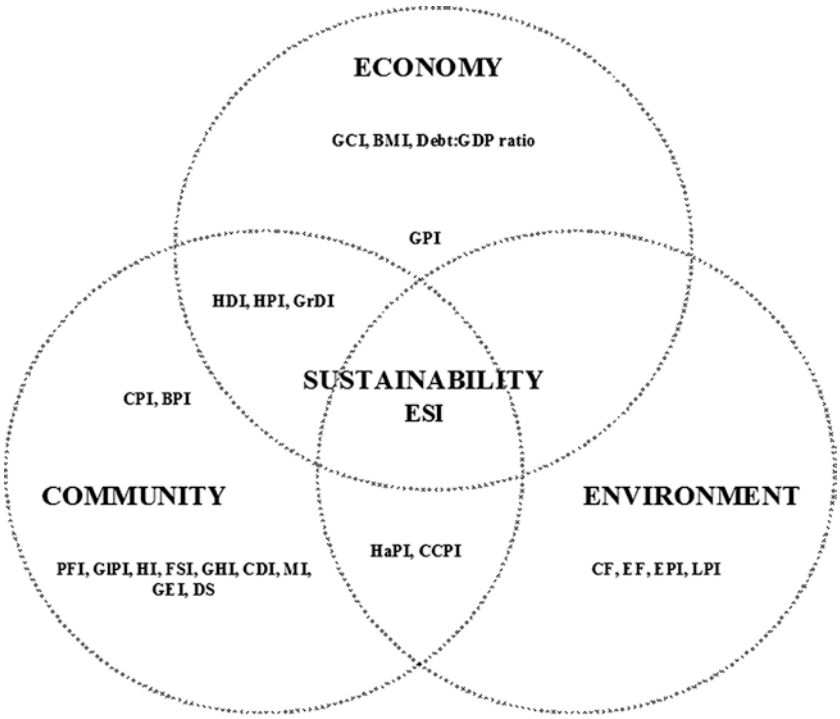

Fig. 19.2 The classical conceptual model of sustainable development (exemplified through a classification of the sustainability indicators listed in Table 19.1). Source: Own design

true sustainability—the space where all three spheres interact. Some of the SIs in Table 19.1 rest firmly within one of the spheres, while some sit in spaces where two spheres overlap; the HDI, for example, sits in the overlap between community and economics. The only SI in this group that arguably can be said to at least try and occupy the "sustainability" space in the centre is the Environmental Sustainability Index (ESI), as it has components spanning community, economy and environment, although the Genuine Progress Indicator (GPI) is also close to that point.

In terms of fSIs, the one selected for comparison was the debt-to-GDP ratio, given that it was expected a priori to have the most reporting (at least relative to other EU indicators of fiscal sustainability). This yielded a total of 25 indicators included in the analysis, with the debt-to-GDP ratio shown shaded in Table 19.1. Some of the SIs in Table 19.1 are updated annually, while others are updated on a biannual or even longer basis, but each one selected has a minimum of six values up to 31 December 2012. A further selection criterion was to focus only on those measures that are relevant

globally rather than those that are only relevant at regional levels (e.g. to the EU or the Middle East). Some of the SIs are relatively long-lived (such as the HDI), while others are newer (such as the Global Hunger Index). The final year for inclusion for each index was 2012, and the starting year was the first year that the index was found reported on in at least one newspaper article, a year that might not necessarily correspond with the year in which the index was created.

The number of articles reporting each SI until the end of December 2012 was found using the subscription-based Nexis database and search tools available via the LexisNexis website. The sources selected were "All news, All languages", and at the time of the search, this spanned a total of 6,760 newspapers from countries around the globe. The Nexis database was set to return a list of newspapers and the number of articles per newspaper reporting the index in that year (adjusting for duplicates).

For each SI, the number of articles published each year that mentioned it at least once was summed across all the newspapers. In order to calculate the diversity of reporting, a modification of the Shannon equation designed to measure biodiversity was adopted:

$$H = -\sum_{i=1}^{i=S} p_i \log_2 p_i ,$$

where the symbol $\sum$ means "sum of" (sum over all newspapers from 1 to S), S is the number of newspapers carrying a story on the SI in a year and $\log_2$ means base 2 logarithm. The key variable in the equation is p_i, which represents the proportion of the total sample of newspaper stories mentioning the SI for newspaper i such that

$$p_i = \frac{n_i}{N} ,$$

where n_i is the number of stories mentioning the index in a year for newspaper i and N is the total number of stories mentioning the same index for that year. The greater the value of H (the Diversity Index) is, the greater the diversity of reporting of the SI by newspapers.

Results of the Press Reporting of Indicators

Each SI has a particular pattern of reporting in the newspapers over time. Figures 19.3, 19.4, 19.5 and 19.6 show the patterns for the debt-to-GDP indicator, the HDI, the ESI and carbon footprint (CF), respectively. These

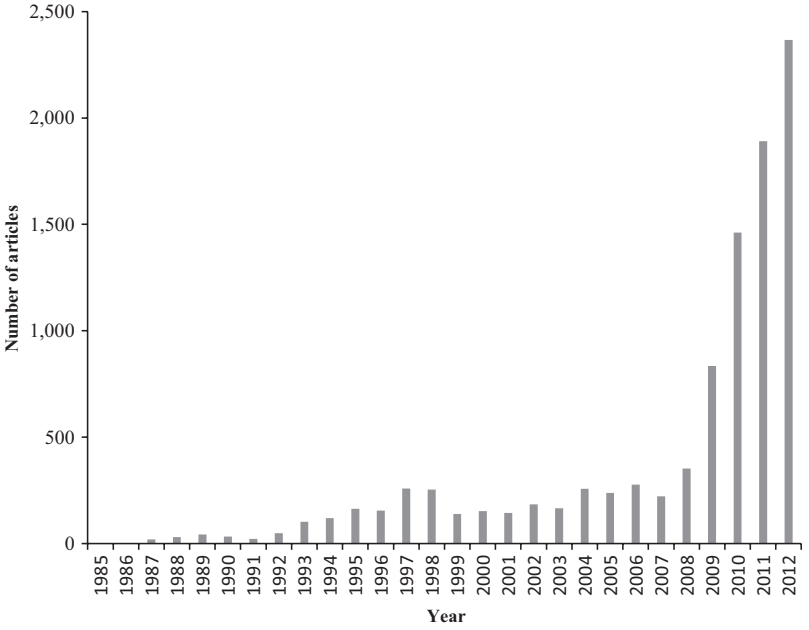

Fig. 19.3 Newspaper reporting of the debt-to-GDP ratio. Source: Own compilation

examples have been chosen to illustrate some of the patterns seen in the article count over time.

For the debt-to-GDP ratio (Fig. 19.3) and the HDI (Fig. 19.4), the trend is a gradual increase in reporting until 2008, followed by something of a surge. In the case of the HDI, there is some evidence that the surge peaked in 2011, while the debt-to-GDP ratio appears to be accelerating until 2012. The surge in reporting of these indices after 2008 may well be related to the economic crisis of that time. Debt-to-GDP became increasingly important, given the scales of debt being accumulated by some developed countries, and the HDI is often associated with "quality of life". Hence, both of these factors—national debt and quality of life—would be expected to feature strongly in press reporting in that period.

The pattern for the ESI (Fig. 19.5) is very different to those for the debt-to-GDP ratio and the HDI.

There is no surge in reporting after 2008, and the pattern suggests a peak in reporting in 2005, followed by a decline. Interestingly, the last year of publishing the ESI was 2005, the year that matches the peak in Graph 5. While the ESI was no longer updated after 2005, the fact that it was still being reported in 2012—some seven years later—is intriguing. To some extent, this apparent longevity of the ESI reflects a subtler influence, in that this SI helped

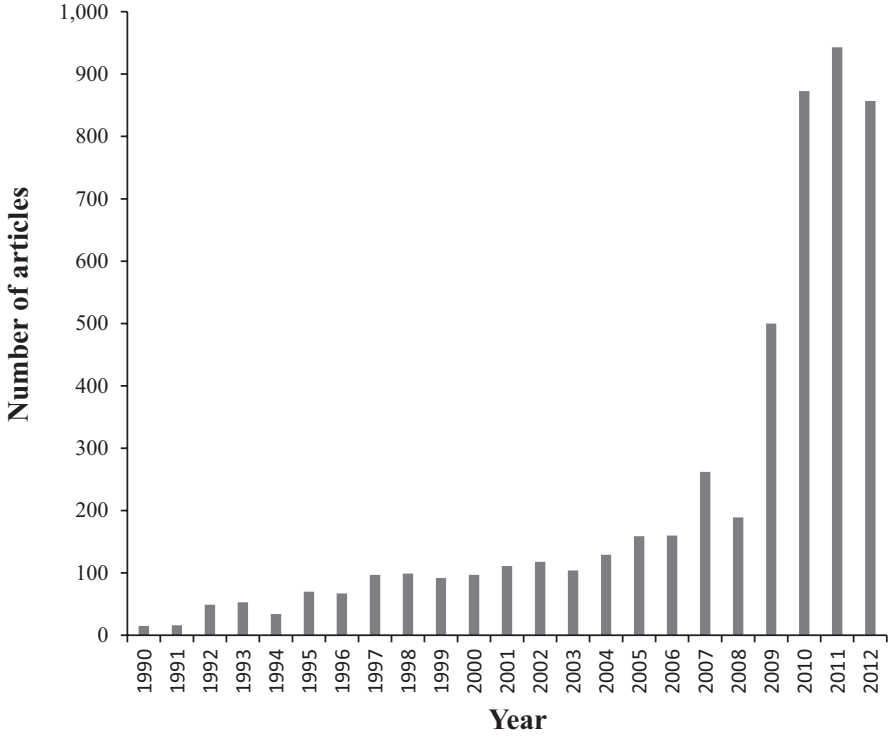

Fig. 19.4 Newspaper reporting of the Human Development Index (HDI). Source: Own compilation

to spawn a number of related indices with similar names. Hence, some of the press reporting of the ESI relates to these offshoots rather than the original index.

Another pattern can be seen for the CF (Fig. 19.6). In this case, there is a sudden surge in reporting in just three years—from 2006 to 2008, after which there is a gradual decline. But the number of articles mentioning the CF is in the tens of thousands—compared with the thousands of the debt-to-GDP ratio, the hundreds of the HDI and the tens of the ESI. These scales are markedly different, and the CF, despite its brief life in terms of reporting, has been reported in far more articles than the other three indices together. Why should this be so? One of the main reasons for the difference in the scale of reporting is related to the transition of the CF from an SI, albeit one that has a number of methodological constructs, to a figure of speech. Thus, the term "carbon footprint" has become associated with a sense of impact arising out of consumption, and it has become a colloquial term in a way that few of the other SIs in Table 19.1 have managed. This provides one of the few examples (the

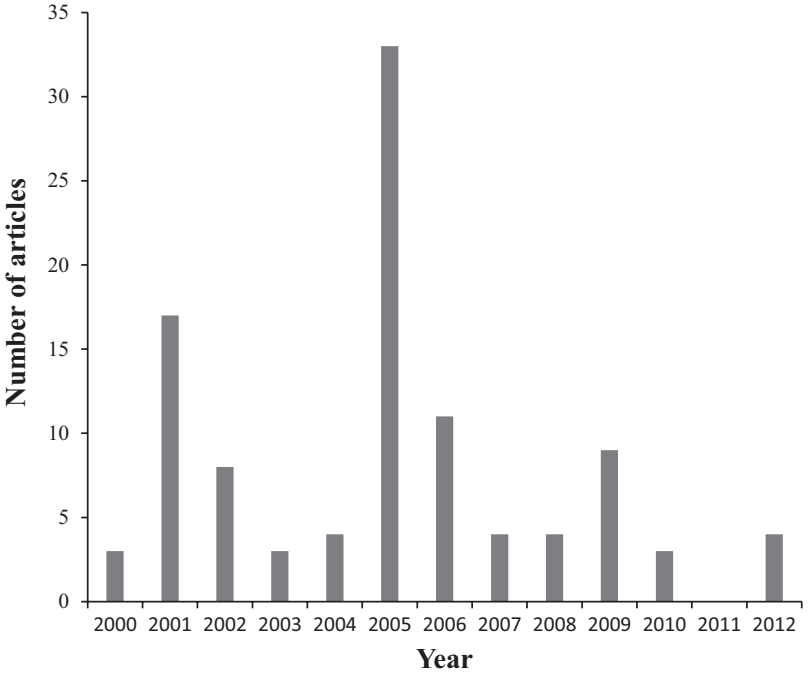

Fig. 19.5 Newspaper reporting of the Environmental Sustainability Index (ESI). Source: Own compilation

EF is perhaps the other) of an SI managing to have an influence within society that goes far beyond that resulting from its measurement—a unique transition from a reporting of numerical values to a conceptualisation of impact. However, the decline in reporting of the CF from the peak of 2008 is intriguing. It perhaps indicates a shift in concern away from the environment towards more economic and development-related issues.

Given that the number of articles per year and the diversity of reporting across newspapers are two ways of assessing the success of SIs, at least in terms of their uptake by newspapers, it is instructive to explore the placement of indices within a two-dimensional space represented by these two measures. Figure 19.7 shows the logarithm of the median number of articles per year on the vertical (dependent) axis and median diversity per year on the horizontal one. Interestingly, these two measures of success have a statistically significant relationship, although this is most apparent if CF is omitted from the analysis, as it is something of an outlier.

Also shown in Fig. 19.7 is an ordinary least squares regression fit (unbroken line) and the upper and lower 95% confidence interval (broken lines). The

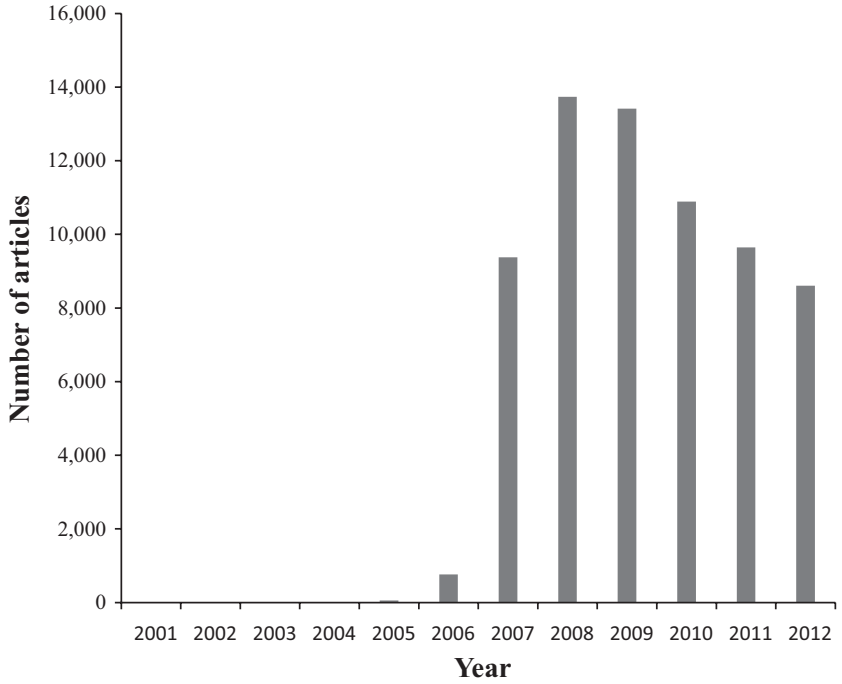

Fig. 19.6 Newspaper reporting for the carbon footprint (CF). Source: Own compilation

regression is statistically significant at $P<0.001$ and suggests that much of the increase in the number of articles is related to a broadening out of the reporting amongst newspapers (that is, as diversity increases, then so does the number of articles). This seems reasonable as one would expect a limit to reporting of an SI within a single newspaper and thus an increase in "use" would be expected to be driven by reporting by a wider range of newspaper titles. Indeed, it may be argued that the successful SIs (those most reported) are on the right-hand side of the graph, while the least successful (least reported) are those towards the left-hand side. The SI having the greatest degree of "success" by far appears to be the CF. The reasons for this have already been discussed. It is also interesting to note that the debt-to-GDP ratio does well in terms of the number of articles but not as well as other SIs in terms of diversity of reporting. Indeed, it is also something of an "outlier" here and sits outside the upper boundaries of the 95% confidence interval. It may well be that the debt-to-GDP ratio is widely reported but in a smaller subset of newspaper titles, essentially the more "serious" titles (the so-called broadsheets) including those that are stronger on financial news. SIs such as the HDI and EF would appear to have a wider appeal across newspaper titles.

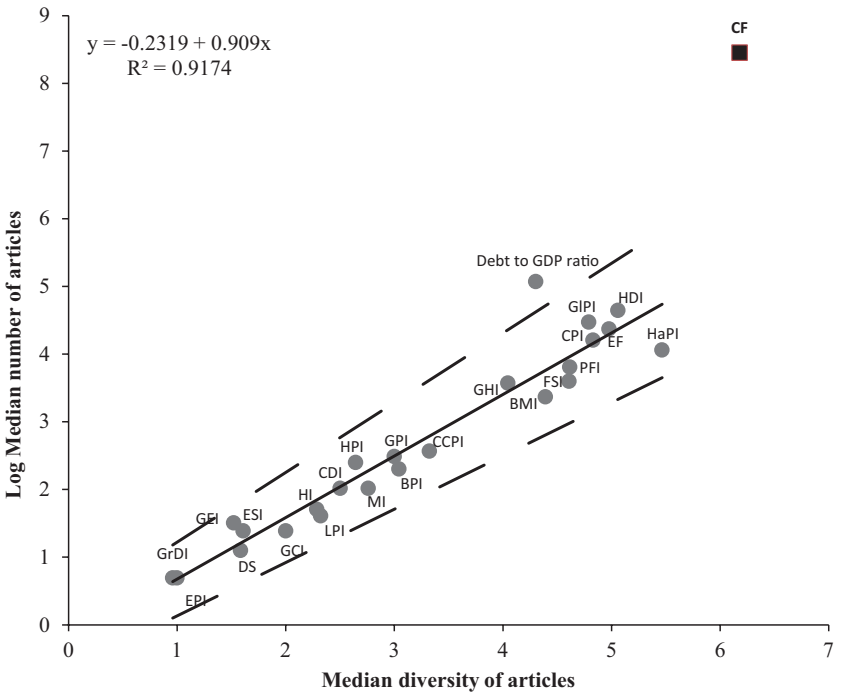

Fig. 19.7 Two aspects of sustainability indicator reporting—number of articles and diversity. The *dashed lines* are the 95% confidence limits (upper and lower) for the regression. Source: Own compilation

Some Lessons for Sustainability Indicator Developers

Drawing out some general lessons from these findings that would help with enhancing the success of SIs is a challenge. In Fig. 19.7, the SIs that stand out are the HDI, CPI and the footprint indices. The popularity of the footprint indices is perhaps related to their being very adaptable (in methodological terms) and also that their names have become synonymous with a general sense of environment impact. The HDI is certainly one of the most strongly promoted indices in the sense that new versions have been released each year since 1990 with accompanying fanfare from the UNDP. The HDI is an index that attempts to assess a very broad concept—human development—and this has, in turn, become synonymous with "quality of life" in the eyes of many, and hence may have some attraction for the newspaper readership in developed countries. Indeed, are the "booms" in the reporting of the HDI and debt-to-GDP ratios in Figs. 19.3 and 19.4 related? Are they both being

employed to illustrate the impact of the financial crisis, albeit from different perspectives? The timing of the surge in reporting—2008—would suggest that this is indeed the case. Other SIs that have done well in terms of at least one of the dimensions of success in Fig. 19.7 are the Happy Planet Index (HaPI) in terms of diversity of reporting and the debt-to-GDP ratio in terms of the number of articles.

The success of some SIs in Fig. 19.7 contrasts markedly with the relative lack of success of others. The relatively low reporting of the "Democracy Score" is perhaps understandable, as this is specifically focussed on the states of the old Soviet Union, and thus may perhaps be regarded as quite specialised. Two of the SIs from the Human Development Reports (HDRs)—the Gender Development Index (GDI) and the Gender Empowerment Measure (GEM)—have relatively low "success" as indeed do the ESI and its relative, the EPI. Both the GDI and GEM are focussed on gender equality, and this has been a major theme in development and indeed has been a centre of many national policies. However, they are well behind the HDI and the other index contained within the HDRs—the Human Poverty Index (HPI). The low "success" of the ESI and the Environmental Performance Index (EPI) is perhaps related to their being quite complex constructs with many components and focussed primarily on the environment. Hence, they may appear to be more abstract in terms of what they "capture" than the two footprint indices.

The pattern in Fig. 19.7 suggests that while the focus of these SIs is clearly important, it is not necessarily the major determinant of success. It is a mistake to generalise, but one would perhaps have expected the Mothers Index, the gender indices and the GPI to have a better success than one focussed on corruption. After all, these indices are broader in nature and covering issues (children, gender equality, etc.) that should have a wider appeal amongst a newspaper readership than something like corruption, which may appear to be quite narrow by comparison. Moreover, it is not necessarily the case that backing from a major international organisation guarantees success, although no doubt "promotion" is a complex element comprising factors such as availability of local personnel to promote the SI, presentation style, use of "league tables" and so on. Neither does age of the SI appear to be a sole requirement for success, although it does seem to be of some influence. There appears to be a special factor which helps explain the success of the footprint SIs, especially the CF, as they have managed to weave their way into language in a way that none of the other SIs have managed to achieve. The reason for this may be partly because of their "adaptability", as many groups and organisations have been able to generate variations on the theme and thus have a sense of

ownership, and also because the term has a resonance in the lay mind (footprint is readily synonymous with impact). Debt has a similar advantage in this regard, as it is something that would be expected to resonate with the readership of newspapers, although when linked to the more opaque concept (at least for many) of GDP, this advantage may become diminished.

It should also be noted that there is nothing here that suggests that an SI necessarily needs to be technically excellent to be successful. Much thought has gone into each of the SIs reported here, successful or otherwise, and their creators go to great lengths to justify their choice of methodology and to use the best quality dataset that they can. All of the SIs in this analysis have certainly had their critics over the years, primarily because there are subjective elements to all of them. It is certainly true that anyone creating an SI or modifying an existing one has a duty of transparency—to set out the methodology and assumptions in a way that others can follow—but it should be noted that acceptance of all of these by other experts in the field may be mixed. This is true for even the most successful of the SIs—such as the HDI and the CF. In addition, of course, there is the matter of how the SIs are used. In this chapter, "use" has admittedly been taken to mean within the very narrow context of newspaper reporting, but even here, there can be divergent views on whether the SI has been reported in the "right" way. An example is provided by the HDI, an SI often equated with "quality of life" but which was actually designed to be a measure of human development; these terms are not synonymous. Similarly, one could dispute the use of the debt-to-GDP ratio as a measure of fiscal sustainability and also make a case that this helps provide an unwise focus on monetary flow within an economy as the most important aspect that needs to be considered. For example, consider the following quotation from the EU:

> GDP has also come to be regarded as a proxy indicator for overall societal development and progress in general. However, by design and purpose, it cannot be relied upon to inform policy debates on all issues. Critically, GDP does not measure environmental sustainability or social inclusion and these limitations need to be taken into account when using it in policy analysis and debates. (Commission of the European Communities 2009, p. 2)

This statement is patently true, of course, as GDP was not designed to measure social inclusion or environmental sustainability. Indeed, at one level, it seems highly unfair to criticise an SI for not doing what it was never intended to do. Nonetheless, the "Beyond GDP" debate has tried to extend the consideration of SIs beyond those that were designed to measure economic

performance. However, while framed in terms of SIs, this is not a debate about indicators per se but the competing visions of what can be regarded desirable, of which they are but expressions; at the end of the day, it is the ideas that matter and not the indicators. Therefore, the use of the debt-to-GDP ratio by newspapers is a symptom of the popularity of the idea that national debt is an important dimension for discussion. Even so, SIs can also be said to be promoters of an idea as they raise awareness. Being able to provide a measure of something is a powerful aid to expression, and while the debt-to-GDP ratio is a reflection of a perception that debt is important, it can also help crystallise that concept in ways that, for example, allow nation states to be compared, and thus help to reinforce the idea it is meant to reflect. This two-way relationship between ideas and their SIs is often underplayed.

Given all the above, it is clear that SIs are dynamic beasts; they are born, live and die. They reflect as well as reinforce. They even evolve according to pressures placed on their creators, and sometimes their death occurs because other SIs take their place. This interplay of SIs, indeed the "sustainability of SIs", has been underexplored in the literature, but would provide an interesting space for research.

Conclusion: A New Dimension to Sustainability Indicator Development

An important point to make here is that the users of an SI, and here it is newspaper reporters, will not necessarily be familiar with the technical details resting behind the SI or indeed make any effort to seek them out. Here rests something of a conundrum. The raison d'être of all SIs (fiscal or otherwise) is that they are created to allow non-specialists to digest and use what can be complex data and ideas. They are thus presented as de facto "black boxes" and, understandably, their consumers treat them in that way and take it on trust that they "work". This, of course, puts a great deal of responsibility on the shoulders of the SI creators, but given that these tools exist to be used, it may seem odd that this aspect has received so little attention by researchers. To date, the attention of SI developers has been almost entirely on technical excellence defined in terms of getting the components and weightings "right", together with justifying any data manipulations and making sure that the datasets on which SIs are based are the best available. This, of course, is laudable, but once the SI has been released, there is typically little, if any, monitoring of its use other than to field any criticisms that may be fed back. Given

that the latter tend to also come from other SI technical specialists, the evolutionary process is driven in terms of technical excellence (or at least a perceived excellence). What is perhaps needed is another selection pressure for SI development that takes wider use into account, use by the very people the SI is aimed at—non-specialists. But so little is known about these dynamics, and, in particular, how SIs are used, that there is still much to do.

References

Bandura, R. (2008). *Survey of Composite Indices Measuring Country Performance: 2008 Update* (United Nations Development Programme, Office of Development Studies Working Paper). New York: United Nations Development Programme.

Barabas, J., & Jerit, J. (2009). Estimating the Causal Effects of Media Coverage on Policy-Specific Knowledge. *American Journal of Political Science, 53*(1), 73–89.

Bell, S., & Morse, S. (2008). *Sustainability Indicators. Measuring the Immeasurable* (2nd ed.). London: Earthscan.

Bell, S., & Morse, S. (2011). An Analysis of Factors Influencing Use of Indicators in the European Union. *Local Environment: The International Journal of Justice and Sustainability, 16*(3), 281–302.

Blowers, A., Boersema, J., & Martin, A. (2012). Is Sustainable Development Sustainable? *Journal of Integrative Environmental Sciences, 9*(1), 1–8.

Commission of the European Communities. (2009). *Communication from the Commission to the Council and the European Parliament: GDP and Beyond – Measuring Progress in a Changing World*. Brussels: Commission of the European Communities.

Dahl, A. L. (2012). Achievements and Gaps in Indicators for Sustainability. *Ecological Indicators, 17*, 14–19.

Frederiksen, F., & Gudmundsson, H. (2013). Policy Use and Influence of Indicators. *Ecological Indicators, 35*, 1–2.

Guinee, J. B., Heijungs, R., Huppes, G., Zamagni, A., Masoni, P., Buonamici, R., Ekvall, T., & Rydberg, T. (2011). Life Cycle Assessment: Past, Present, and Future. *Environmental Science and Technology, 45*(1), 90–96.

Herzi, A. A. (2004). Sustainability Indicators System and Policy Process in Malaysia: A Framework for Utilisation and Learning. *Journal of Environmental Management, 73*(4), 357–371.

Holt, D., & Barkemeyer, R. (2012). Media Coverage of Sustainable Development Issues – Attention Cycles or Punctuated Equilibrium? *Sustainable Development, 20*(1), 1–17.

Jeswani, H. K., Azapagic, A., Schepelmann, P., & Ritthoff, M. (2010). Options for Broadening and Deepening the LCA Approaches. *Journal of Cleaner Production, 18*(2), 120–127.

Leeuw, S., Wiek, A., Harlow, J., & Buizer, J. (2012). How Much Time Do We Have? Urgency and Rhetoric in Sustainability Science. *Sustainability Science, 7*(Supplement 1), 115–120.

Mekelberg, D. (2012). The Influences of Global News Messages on National Governability: An Environmental Case Study. *Israel Affairs, 18*(2), 268–285.

Morse, S. (2004). *Indices and Indicators in Development. An Unhealthy Obsession with Numbers*. London: Earthscan.

Morse, S. (2011a). Harnessing the Power of the Press with Indices. *Ecological Indicators, 11*(6), 1681–1688.

Morse, S. (2011b). Attracting Attention for the Cause. The Reporting of Three Indices in the UK National Press. *Social Indicators Research, 101*(1), 17–35.

Morse, S. (2013). Bottom Rail on Top: The Shifting Sands of Sustainable Development Indicators as Tools to Assess Progress. *Sustainability, 5*(6), 2421–2441.

Morse, S. (2014). Stirring the Pot. Influence of Changes in Methodology of the Human Development Index on Reporting by the Press. *Ecological Indicators, 45*, 245–254.

Pissourios, I. A. (2013). An Interdisciplinary Study on Indicators: A Comparative Review of Quality-of-life, Macroeconomic, Environmental, Welfare and Sustainability Indicators. *Ecological Indicators, 34*, 420–427.

Policy Influence of Indicators (POINT). (2008–2011). [Website]. Retrieved from http://cordis.europa.eu/project/rcn/89898_en.html

Rametsteiner, E., Pülzl, H., Alkan-Olsson, J., & Frederiksen, P. (2011). Sustainability Indicator Development – Science or Political Negotiation? *Ecological Indicators, 11*(1), 61–70.

Schmidt, A., Ivanova, A., & Schafer, M. S. (2013). Media Attention for Climate Change Around the World: A Comparative Analysis of Newspaper Coverage in 27 Countries. *Global Environmental Change, 23*(5), 1233–1248.

Turcu, C. (2013). Re-thinking Sustainability Indicators: Local Perspectives of Urban Sustainability. *Journal of Environmental Planning and Management, 56*(5), 695–719.

20

Sustainability of Public Debt: A Dangerous Obsession?

Christophe Blot

Introduction

Public debt has significantly risen since 2007, reaching high levels for peacetime. From 2007 to 2015, the average increase in the debt of OECD (Organisation for Economic Cooperation and Development) countries has exceeded 40 GDP (gross domestic product) points (Fig. 20.1), with record increases for Ireland (92 points) and Spain (77 points). The debt-to-GDP ratio has increased for all OECD countries except Israel, Norway and Switzerland, reviving fears of public default in advanced countries. These concerns were particularly acute in the euro area after the outbreak of the Greek crisis. Contagion reached other countries on the periphery (Ireland, Portugal, Spain and Italy) for which a risk of unsustainable fiscal policy was presumed. Although the public finances of these countries showed some heterogeneity—in terms of level of public debt or size of public deficit—sovereign yields relative to Germany increased sharply and did not recede until Mario Draghi's statement according to which the European Central Bank (ECB) stood ready to do "whatever it takes" to save the euro. The concepts of fiscal sustainability and fiscal space then resurfaced and illustrated the need to resort to qualitative and quantitative indicators to assess the soundness of public finance (Ghosh et al. 2013; Berrittella and Zhang 2015).

C. Blot (✉)
Observatoire français des conjonctures économiques (OFCE) Sciences Po, Paris, France

D.V. Malito et al. (eds.), *The Palgrave Handbook of Indicators in Global Governance*,
https://doi.org/10.1007/978-3-319-62707-6_20

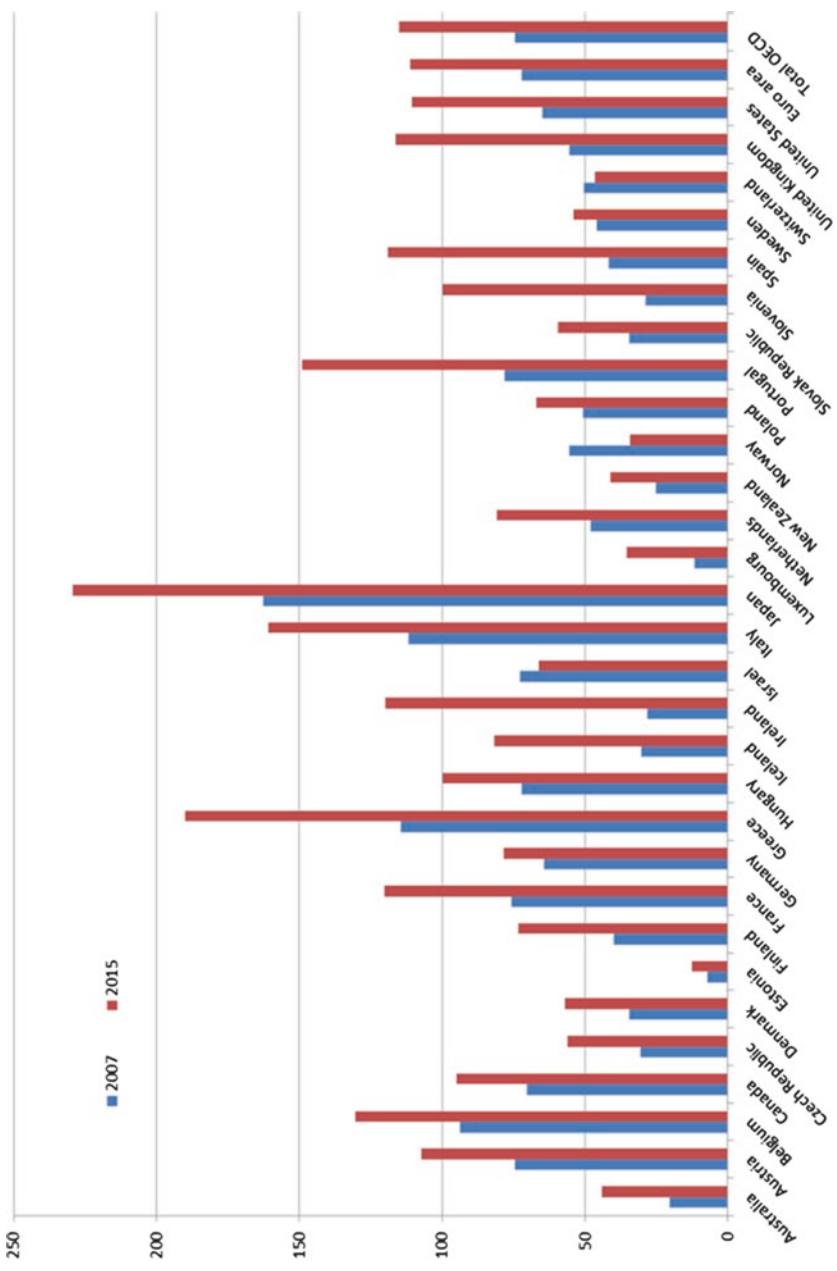

Fig. 20.1 General government gross financial liabilities (percentage of GDP). Source: Own compilation based on OECD, EO 98 database

Sustainability is generally considered to be a long-term issue, and it concerns understanding whether a situation or a policy can be maintained without damaging or threatening the situation in the long term. It is intrinsically a tricky issue since it is related to risks that may only materialise in the future. Assessing sustainability, therefore, involves expectations of future contingencies. When the concern is fiscal sustainability, it is based on the government's intertemporal budget constraint (IBC). Fiscal policy is therefore said to be sustainable when current debt is backed by expected flows of fiscal revenues and expenditures. However, such an approach may be of little interest in providing rapid and clear information on fiscal sustainability. There is therefore a need to build simple, easy-to-compute and immediately available indicators. However, these indicators must be reliable. Otherwise, there is a risk of providing misleading policy recommendations, which would not only lead to missing the initial objective but would also have negative economic or social consequences. For example, if indicators fail to detect an unsustainable policy, the government may default on debt, raising financial instability. Thus, it is not only crucial for indicators to send the right signal but also for them to send it in advance to avoid unsustainability ending in a sovereign debt crisis, triggering a public default. The history of public default has particularly emphasised long resolution times for such crises (Oosterlinck 2013) and then pointed to the difficulty for the defaulter in having access to financial market funding. Conversely, if indicators of fiscal sustainability provide the wrong information, signalling that a policy is unsustainable when it is not, they may force the government to change policy and either raise taxes or cut public expenditure. Fiscal consolidation may then reduce the scope for macroeconomic stabilisation, disrupt redistribution policy and force the government to reduce the supply of public goods and services.

During the 1990s, Blanchard et al. (1990) proposed simple indicators that are able to not only show whether fiscal policy is sustainable or not but also quantify the required change in policy. However, these indicators also turn on a long-term analysis and consequently fail to account for the risk of liquidity. Governments may be forced to default due to phenomena of self-fulfilling prophecies in the financial markets. Fiscal policy becomes unsustainable not because of a risk of long-term insolvency but because bondholders expect that a default will occur and start to turn away from sovereign debt. The government rapidly becomes unable to service interest payments or meet its outlay commitments if it is unable to raise funds on the markets. Moreover, fiscal sustainability indicators crucially hinge on strong hypotheses regarding interest rate and growth. These variables are assumed to be constant and exogenous, whereas in practice they are not. Notably, they are strongly interconnected

with fiscal policy variables. Failing to account for feedback effects implies biased judgements on fiscal sustainability. This deficiency is inherently linked to the partial equilibrium setting used to assess fiscal sustainability. The situation in terms of public finance is not independent of the macroeconomic situation, and vice versa. Accounting for economic interdependencies also highlights potential conflicts between different economic and social objectives. Fiscal sustainability matters, but growth, employment, financial stability and social welfare are at least as important as fiscal sustainability, and the only way to address the conflict of objectives is to be aware of the potential trade-offs.

The aim of this chapter is to provide a critical analysis of fiscal sustainability indicators and to show that public sustainability should not be considered independently of other economic and social variables. To that end, section "Assessing Fiscal Sustainability: From Theory to Practical Indicators" first defines the concept of sustainability applied to fiscal policy and presents the most common indicators to analyse fiscal sustainability. Section "Fiscal Sustainability: Don't Lose Sight of Liquidity" shows that liquidity also matters when judging sustainability, while section "Dealing with Public Debt Unsustainability: The Endogeneity Problems" provides a critical view highlighting the endogeneity of the main hypotheses used to compute fiscal sustainability indicators. Finally, section "Conclusion" concludes that assessing fiscal sustainability inherently remains a tricky issue and that reducing public debt in the recent crisis period has led policymakers to overlook other objectives, such as growth and employment.

Assessing Fiscal Sustainability: From Theory to Practical Indicators

Sustainability refers to the ability of a policy to be maintained over time. When policy is unsustainable, it is expected to change at some point in the future. Considering this simple definition, the role of sustainability indicators is precisely to signal unsustainability. It also aims to quantify policy change when it is needed. Considering public finances, sustainability is strongly related to the solvency of the government. Will it be able to maintain its current policy, namely the current path of public expenditure and fiscal receipts, without threatening its future solvency? If it is not changed, an unsustainable fiscal policy may lead to a future default. Solvency is generally assessed through the government's IBC. We first explore the theoretical foundations of sustainability indicators and then describe the most common indicators and

the key concepts generally used when fiscal sustainability is analysed (Chalk and Hemming 2000; Balassone and Franco 2000; Giammarioli et al. 2007).

Theoretical Foundations of Sustainability Indicators

The concept of government debt sustainability refers to the ability of the government to remain solvent, given the current level of debt and the future dynamics of public expenditure and revenue. It is intrinsically related to the debt dynamic equation. In each period, nominal public debt increases with nominal primary expenditure[1] and the interest payments on current debt and decreases with nominal receipts:

$$B_t - B_{t-1} = G_t - T_t + i.B_{t-1.} \tag{20.1}$$

The dynamic of nominal public debt is of only limited interest per se and should be scaled according to the country's size. Sustainability is then assessed using public debt measured as a ratio to GDP. Dividing equation (20.1) by nominal GDP (Y_t) yields:

$$b_t = g_t - \tau_t + \left(\frac{1+i}{1+\gamma}\right).\ b_{t-1,} \tag{20.2}$$

where $b_t = \frac{B_t}{Y_t}$ is the debt-to-GDP ratio at time (t), $g_t = \frac{G_t}{Y_t}$ is the ratio of primary expenditure to GDP and $\tau_t = \frac{T_t}{Y_t}$ is the ratio of total receipts. The nominal interest rate (i) and the GDP growth rate (γ) are assumed to be exogenous and constant. As assessment of solvency is based on long-term prospects, the interest rate and the growth rate in equation (20.2) are assumed to be equal to their long-term values.

By rearranging equation (20.2), current debt can be expressed as a function of future debt, future primary expenditures and receipts:

$$b_t = \left(\frac{1}{\rho}\right)\left(\tau_{t+1} - g_{t+1}\right) + \left(\frac{1}{\rho}\right).b_{t+1,} \tag{20.3}$$

where $\rho = \frac{1+i}{1+\gamma}$. Hereafter, we will suppose that the critical gap (the difference between the interest rate and the GDP growth rate) is positive: $(i - \gamma) > 0$.

Solving equation (20.3) forward, gives:

$$b_t = \sum_{n}^{j=1} \left(\frac{1}{\rho}\right)^j \left(\tau_{t+j} - g_{t+j}\right) + \left(\frac{1}{\rho}\right)^n . b_{t+n.} \tag{20.4}$$

According to equation (20.4), the present value of the debt-to-GDP ratio is equal to the expected present value of the future primary balance plus the present value of future debt. Public debt is sustainable if and only if the following condition holds:

$$\begin{aligned} &\underbrace{\lim}_{n\to\infty} \left(\frac{1}{\rho}\right)^n . b_{t+n} = 0 \Leftrightarrow \underbrace{\lim}_{n\to\infty} \left(\frac{1}{\rho}\right)^n . \frac{B_{t+n}}{Y_{t+n}} = 0 \\ &\Leftrightarrow \underbrace{\lim}_{n\to\infty} \left(\frac{1+\gamma}{1+i}\right)^n . \frac{(1+\kappa)^n B_t}{(1+\gamma)^n Y_t} = 0 \end{aligned} \tag{20.5}$$

$$\Rightarrow b_t = \sum_{\infty}^{j=1} \left(\frac{1}{\rho}\right)^j \left(\tau_{t+j} - g_{t+j}\right). \tag{20.6}$$

Equation (20.5) is the transversality condition, which requires that the present value of debt at an infinite horizon converges to zero. It simply states that the government cannot run a Ponzi game on debt.[2] It holds when the growth rate of the public debt ratio does not exceed ρ, which is equivalent to saying that the growth rate of nominal debt (κ) should not exceed the interest rate. Equation (20.6) describes the government's IBC. The government is solvent—that is, current policy is sustainable—if the current level of debt is repaid with the expected present value of future primary surpluses. It should be noted that the government does not have to run permanent surpluses but that the current level of debt should be backed by large enough primary surpluses. Otherwise, public debt is unsustainable and fiscal policy has to be changed. The IBC emphasises that fiscal sustainability is a long-term issue, as it hinges on expected future primary surpluses from now to infinity. Moreover, it should be noted that sustainability also depends on the critical gap, which is assumed to be exogenous. Subsequently, the problems raised by these hypotheses is discussed.

Starting from this definition, a large empirical literature has developed—for a large set of countries and using different econometric techniques—to test whether condition (20.5) and more often condition (20.6) holds. Arguing

that condition (20.5) holds when there is positive feedback of public debt to the primary surplus, Bohn (1998) suggested estimating fiscal policy rules (Bohn 1995, 2007)—the reaction of the primary surplus to several economic variables and to past public debt—to gauge fiscal sustainability. This literature remains mainly technical and has not provided clear conclusions. Moreover, sustainability is generally inferred over a long period, whereas there may be periods over which public debt is sustainable and others where it is not. Efforts may be made so that debt becomes sustainable at the end of the sample, but this is not captured by tests based on the average behaviour over the period. To overcome this limit, Davig (2005) and Doi et al. (2011) propose using a regime-switching approach to test for the existence of different fiscal regimes—a sustainability regime or an unsustainability regime—over time. Using the same regime-switching approach, Aldama and Creel (2016) suggest disentangling short-term and long-term sustainability and show that this distinction is needed to take into account the duration of fiscal regimes in assessing fiscal sustainability. Nevertheless, the fact remains that econometric tests do not provide direct indicators for assessing fiscal sustainability.

An alternative to formal testing would be to assess sustainability by calculating the right-hand side of equation (20.6) on the basis of forecasts of future (to infinity) expenditure and fiscal receipts. It should, however, be admitted that forecasting developments in the main fiscal variables over the long term would inevitably lead to large approximations. Moreover, Blanchard et al. (1990) claim that if equation (20.6) is not satisfied, it would only show that some change needs to be implemented at some point in the future. When the policy change should be implemented is not clear and the size of the policy change remains subject to considerable approximations. Therefore, in parallel to formal testing, economists have also provided simple and easy-to-compute indicators to assess fiscal sustainability and quantify the change in policy needed (either through taxes or public expenditure).

Sustainability Indicators in Practice

In order to assess fiscal sustainability, it is useful to resort to simple indicators which are rapidly available and easy to compute. The aim of these indicators is to provide a clear message regarding sustainability (is a change in policy required?) and to quantify the change in policy (if so, what would be the size of the adjustment?). Blanchard et al. (1990) suggest that fiscal policy is sustainable when "the debt-to-GDP ratio eventually converges back to its initial level" at some point in the future. It can be shown that this definition is

compatible with the IBC. Starting from the IBC (20.6), we calculate the constant level of the primary deficit ($\tau^* - g^*$) such that the debt ratio remains constant:

$$\left(\tau^* - g^*\right) = \frac{i-\gamma}{1+\gamma}.b_{t.} \tag{20.7}$$

It is then easy to quantify the fiscal effort needed to stabilise public debt as the gap between ($\tau^* - g^*$) and ($\tau_t - g_t$). For ($\tau^* - g^*$) < ($\tau_t - g_t$), the government needs to increase its primary balance, that is, increase taxes or cut public expenditure. An alternative version of the indicator is to compute the tax gap (or the public spending gap) as the difference between the constant tax rate (or the constant public spending ratio) that stabilises debt, given the current level of public expenditure (or the ratio of fiscal receipts to GDP), the interest rate and the GDP growth rate (Blanchard 1990). For the tax gap, we write:

$$\left(\tau^* - \tau_t\right) = g_t - \tau_t + \frac{i-\gamma}{1+\gamma}.b_{t.} \tag{20.8}$$

This indicator is easy to calculate and provides clear interpretations. For ($\tau^* - \tau_t$) > 0 or equivalently ($g^* - g_t$) < 0, the tax rate should be increased to stabilise the debt-to-GDP ratio, or, equivalently, the level of public spending should be reduced to make fiscal policy sustainable. The size of the tax gap provides information about the efforts needed to achieve sustainability.

A first limit of this indicator is that it does not take into account the situation of the economy—whether it is in recession or in a boom period. However, public spending (or fiscal revenues) increases when the output gap[3] is negative, so fiscal sustainability measured using equation (20.8) automatically deteriorates in bad times. In bad times, fiscal receipts are lowered and social expenditure increases. The indicator calculated using equation (20.8) therefore provides a biased view of the fiscal position and indicates that efforts should be made to improve the primary balance, whereas fiscal sustainability would improve with the recovery. Equation (20.8) can be modified to take account of the effect of the cyclical position of the economy on fiscal variables. To this end, the indicator can be calculated with cyclically adjusted values for expenditure $\left(\widetilde{g}\right)$ and revenue $\left(\widetilde{\tau}\right)$, that is, values for expenditure and revenue which do not move with the cyclical position of the economy[4]:

$$\left(\tau^* - \tilde{t}_t\right) = \tilde{g}_t - \tau_t + \frac{i-\gamma}{1+\gamma}.b_{t.} \tag{20.9}$$

The indicator can also be modified to address the issue of sustainability at a given horizon. It is also easy to target another value for the public debt to comply with institutional arrangements such as the fiscal compact in the euro area. Considering that the level of the debt ratio should converge to its initial value in n periods given the expected path for public outlays, we can compute the constant ratio satisfying this constraint by modifying equation (20.4). Thus, fiscal policy is sustainable if $\tau^* \geq \tau$ with τ^* such that:

$$\tau^* = \frac{i-\gamma}{1+\gamma}.\left[b_t + \frac{\rho^n}{1-\rho^n}.\sum_{n}^{j=1}\left(\frac{1}{\rho}\right)^j.g_{t+j}\right]. \tag{20.10}$$

Expression (20.10) becomes more complicated due to its forward-looking nature but is based on the same philosophy: seeking the tax ratio at which public debt is sustainable at a given horizon. As stressed before, it relies on the ability to have available projections for public spending. Nevertheless, this indicator enables future commitments and liabilities whose payments fall due in the future to be taken into account. For instance, implicit or explicit commitments regarding the payment of pensions in public systems may weigh on future public finances and impact fiscal sustainability (Doi et al. 2011).[5] More generally, taking into account demographic growth and its consequences for—social—public spending (European Commission 2015) would also change the assessment of fiscal sustainability (see Balassone et al. 2009). Although demographic projections are supposed to be reliable, one has not only to take into account future spending but also resources that will be dedicated to social spending. Concerning public pension schemes, assessing sustainability requires taking into account liabilities (payment of future pensions) and assets (contributions), which relies on forecasting the dynamics of employment and wages (Balassone and Franco 2000). Additional hypotheses are therefore needed to compute the indicator, with the risk of making it less reliable if it hinges on uncertain scenarios for future social spending and future wage developments.

Although a constant debt ratio is compatible with the non-Ponzi game condition, there is no reason to consider that the debt ratio should converge back to its initial value. Any finite value of the nominal stock of public debt is compatible with equation (20.5) for ($i > \gamma$) and $n \to \infty$. Why should the

public debt ratio be stabilised around 60% and not 80%? This threshold is certainly country-specific and may explain why fear of default has not arisen in Japan despite a debt-to-GDP ratio exceeding 225%, whereas it has been a major concern for Greece with public debt below 200%. Condition (20.5) holds under much more general conditions than the initially suggested requirement to stabilise the debt ratio at its initial value. In fact, neither conditions for fiscal sustainability nor fiscal sustainability indicators provide theoretical foundations for the optimal or the maximum ratio for public debt. The optimal level for public debt remains an unexplored issue. More generally, this shows that assessing fiscal sustainability in a partial equilibrium has significant flaws which limit the relevance of the indicators in (20.7), (20.8), (20.9) and (20.10). The sustainable level of debt also crucially depends on the demand for a public asset, an issue that can only be addressed in a more general model (Bohn 1995, 2008).[6]

Although it does not circumvent the argument sketched above, one may consider that public debt is sustainable when it converges to a target $\left(\bar{b}\right)$. The tax rate compatible with the target can be derived from equation (20.4):

$$\tau^{*} = g + \left(\frac{i-\gamma}{1+\gamma}\right).\left(\frac{(1+i)^{n}}{(1+i)^{n}-(1+\gamma)^{n}}\right)b_{t} - \left(\frac{i-\gamma}{1+\gamma}\right).\left(\frac{(1+\gamma)^{n}}{(1+i)^{n}-(1+\gamma)^{n}}\right).\bar{b}. \tag{20.11}$$

For $\tau^{*} > \tau$, the tax rate must be increased in order to meet the target. The effort increases for a higher inherited debt b_t and for a lower target $\bar{b}$. The target for public debt may result from institutional arrangements. According to the existing fiscal rules,[7] the public debt of EU member states should not exceed 60% of GDP or should at least diminish sufficiently towards the 60% threshold. The 60% debt ratio has no theoretical foundations and may be interpreted as an institutional constraint that some EU countries have mutually agreed to comply with (Blot et al. 2014a).

Beyond the reliability of hypotheses regarding the future path for fiscal variables (public spending when the indicator is expressed as a tax gap), some technical issues are worth mentioning. Assessing fiscal sustainability requires precisely defining the scope of the sectors of reference which are included, and deciding whether the definition is applied to gross or net debt. Computing an indicator only for central government and overlooking local government entities and public social administrations would clearly lead to missing the

objective of the indicator, as it would not take into account fiscal developments that have direct consequences for public debt. It is therefore clear that sustainability indicators should be applied to all government entities: the central government, local governments and social administrations. There is yet more uncertainty concerning public enterprises. Governments sometimes provide guarantees to firms in the private sector and to financial institutions. This was notably the case during the global financial crisis. The occurrence of a banking crisis would then significantly modify fiscal sustainability, as was illustrated by Ireland and Spain (see Wheelan 2014; or McHale 2012). Even when public guarantees have not been explicitly granted to financial institutions, the "too big to fail" doctrine involves implicit contingent liabilities for governments, which may increase the risk of unsustainability or the perceived risk of unsustainability. However, taking into account these contingent liabilities in a sustainability indicator is a difficult task. It implies estimating the probability of a banking crisis and assessing the expected cost of the crisis. Laeven and Valencia (2012) show that these crises have huge fiscal cost (6.8 points of GDP on average for all countries over the period 1970–2011). Nevertheless, this estimate is not country-specific. Although the risk of banking crises has increased since the 1970s, they remain rare events at the national level, so there is no reliable assessment of a country-specific expected fiscal cost.

Finally, although statistical information on public debt most often refers to gross figures, assets held by the government should also certainly matter when assessing sustainability. Indeed, the government always has the opportunity to sell assets to reduce or repay its debt. However, data on assets are less precise than data on liabilities. The value of real assets is notably subject to higher uncertainty. Statistics on net debt generally cover gross debt less financial assets. Moreover, the liquidity of some public assets—the ability of assets to be sold rapidly without losing value—may be weak. Consequently, for practical reasons, fiscal sustainability indicators are generally computed for gross debt (Balassone and Franco 2000). However, as stressed above, the background for fiscal sustainability is related to solvency. Therefore, this would be a long-term issue, so that the liquidity properties of government assets should not be a concern unless liquidity issues also influence fiscal soundness.

Fiscal Sustainability: Don't Lose Sight of Liquidity

As stressed above, the background for fiscal sustainability indicators is found in the IBC and refers to solvency issues. It is intrinsically related to expectations regarding the dynamics of the primary fiscal balance. However, Bohn

(2008, p. 12) reminds us that an "agent's ability to borrow is constrained by other agents' willingness to lend", suggesting that even when the fundamentals are safe, governments can face difficulties if for any exogenous reason no one is ready to finance the deficit. Market sentiment becomes critical in assessing sustainability, and a government may face difficulty at any point in time in meeting its commitments—servicing debt and paying expenditures, which fall due even if the IBC is satisfied. Consequently, sustainability "requires the government to be both solvent and liquid" (ECB 2012, p. 59). In fact, the issue of liquidity problems is related to the hypothesis of a constant interest rate. When the risk or the perceived risk of default increases, the interest rate rises and the cost of servicing debt increases. In the extreme case, no one is willing to lend to the government, which will be forced to default if it is not able to pay interest or its current outlays. In practice, liquidity and solvency problems are hard to disentangle, as default generally occurs because of the inability of the government to satisfy its financing needs, that is, because of illiquidity. Causality between solvency and liquidity problems is generally not clear. Does a liquidity squeeze arise because of the government's insolvency—financial markets realise that the IBC will not be met and that default will eventually occur—or is insolvency triggered by a lack of liquidity and market sentiment?

Considering an economy with perfect foresight, perfect information and efficient financial markets, insolvency triggers illiquidity. IBC is perfectly forecast by financial markets, and crises occur if and only if current debt is not backed by expected future (from now to infinity) primary surpluses. However, with information frictions, liquidity crises may even occur if the fundamentals are safe. Diamond and Dybvig (1983) first illustrated this phenomenon for the banking industry. Panic is a rational expectations equilibrium if some depositors with imperfect information about the solvency situation of banks start to expect a default and withdraw cash. Then, even depositors who believe that the bank is solvent will find it optimal to withdraw cash, as they are aware that the bank holds illiquid assets and will be unable to meet all the needs of cash. They all have an interest in being served first, which triggers a run on the bank and a default even if the bank is solvent. The model for bank runs emphasises that there may be multiple equilibria: a good equilibrium—without a run—where no one considers that the bank is insolvent and a bad equilibrium where a fraction of depositors expect a default, which triggers a panic. In the bad equilibrium with self-fulfilling prophecies, market sentiment (depositors' sentiment in the banking industry) is crucial for liquidity and hence solvency.

Market sentiments may also be important drivers of interest rates and default in sovereign markets (see Calvo 1988; Cole and Kehoe 1996). This has been illustrated in macroeconomic models where the central analysis departs from the IBC literature and considers that public default results from a political choice. Governments are engaged in a strategic game with financial markets and face a trade-off between consolidating their budgets to reduce public debt, which entails macroeconomic costs, and defaulting on debt, which entails reputation costs and may restrict their future access to market funding. The default option depends not only on fundamentals but also on market sentiment. For low levels of debt, credibility is high, the interest rate paid on sovereign bonds is equal to the risk-free interest rate and default never occurs. Conversely, when the fundamentals are strongly deteriorated, the risk premium on sovereign bonds is high and default always occurs. An intermediate zone exists for the value of fundamentals where credibility is partial (Blot et al. 2016). Within this zone, market expectations are self-fulfilling (Fig. 20.2) and default occurs if and only if it is expected. If the risk premium—and hence the interest rate paid by the government—is high enough, it becomes optimal for the government to default. If the risk premium remains low, the government does not find it optimal to default. The risk premium is not only related to the probability of default but is also driven by risk aversion. In crisis periods, risk-aversion (and conversely the risk premium) may fall (increase), sharply leading investors to seek safe assets and to sell assets considered more risky. The rise in interest rates may create a snowball effect on debt leading to unsustainability.

As illustrated by the IBC (Equation 20.6), the conditions under which debt is sustainable become more stringent for higher interest rates. Rational agents may expect that fiscal policy will become unsustainable if the interest rate keeps on rising. This leads them to sell public assets, which increases the interest rate, making public debt unsustainable. In this situation where the fundamentals are not safe enough but have not deteriorated so much, any bad signal may change the market's perception and lead to a default. In

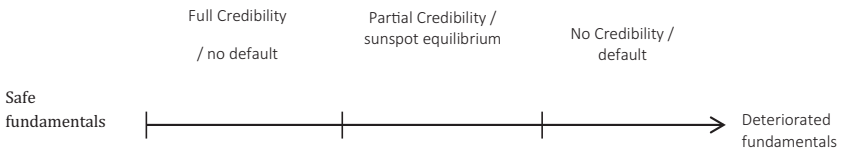

Fig. 20.2 Level of fundamentals and multiple equilibria. Source: Blot et al. 2016

times of crisis, this can be related to risk-aversion or to a negative shock hitting the economy and deteriorating the fiscal balance. As seen during the recent euro-area sovereign debt crisis, the vicious circle can be fuelled by credit rating agencies sending a bad signal and leading the market to distrust the government.

De Grauwe (2012) suggests that such a risk of a liquidity squeeze is higher in monetary unions as governments issue debt in a currency over which they do not have complete control. This would explain why the sovereign debt crisis hit the euro area despite a debt ratio that was close to the US debt-to-GDP ratio. Imperfect control over the value of money issued in the monetary union has two consequences for its members. First, in a monetary union, bondholders have the opportunity to switch from one market (country) to another without being exposed to the exchange rate risk. Spanish and German bonds are both denominated in euros so that market liquidity is the main cost associated with portfolio reallocation. Moreover, in a monetary union, member states alone cannot force the central bank to intervene and play the role of lender of last resort to governments. Conversely, in a stand-alone country, the government has the option to force the central bank to monetise debt, that is, to issue more money to buy bonds on the market. A holder of bonds is aware of this opportunity and will be more reluctant to speculate against the government, as he knows that the central bank may guarantee the price if forced to do so by the government. This was not an option in the euro area before Mario Draghi's statement in July 2012, when he announced the ECB would do "whatever it takes to preserve the Euro. And believe me, it will be enough", which was followed by the Outright Monetary Transactions (OMT) programme through which the ECB stood ready to buy an unlimited quantity of sovereign bonds under certain conditions. Before July 2012, the lack of commitment from the ECB gave power to the financial markets to speculate against sovereign assets. Governments were consequently more prone to a liquidity squeeze.

The rise in sovereign spreads led governments, notably in Spain, Portugal, Greece, Ireland and Italy, to engage in fiscal consolidation policies, which were reinforced to convince financial markets of their credibility to avoid fiscal unsustainability. De Grauwe and Ji (2014) show that there is a positive correlation between sovereign yield spreads and austerity measures. The success of this strategy was, however, limited. In the short term, such strategies entail macroeconomic costs and lower growth (see Section 4 for details). In fact, sovereign spreads declined more significantly after Mario Draghi's speech (Fig. 20.3). Altavilla et al. (2014) confirm the significant impact of the OMT, implemented a few weeks after Mario Draghi's declaration, and find a

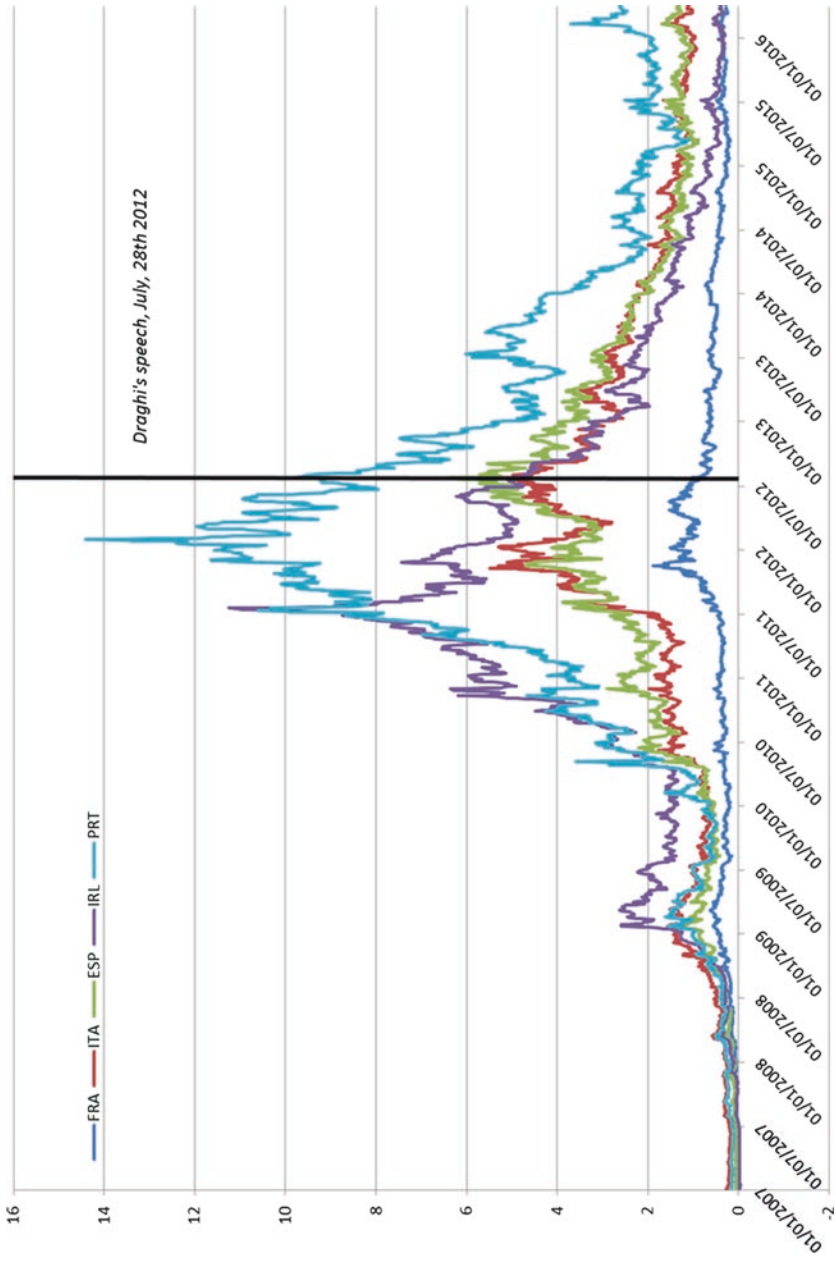

Fig. 20.3 Sovereign spreads in crisis countries (%). Source: Own compilation based on Datastream data

significant decrease in Italian and Spanish government bond yields. In comparison, the German and French bond yields of the same maturity were left unchanged.

These developments suggest that assessing fiscal sustainability with indicators where the interest rate is constant may raise problems, notably in times of financial crisis. These periods are indeed characterised by more financial volatility. Interest rates may change rapidly depending on news, contagion effects and market sentiments. Fundamentals still matter, but the theoretical and empirical literature, which has resurfaced with the euro-area sovereign debt crisis, emphasises that they are not the only drivers of interest rates (Arghyrou and Tsoukalas 2011; Arghyrou and Kontonikas 2012; Bernoth et al. 2012).

Fiscal sustainability indicators thus provide an incomplete picture of the situation and may miss important threats to sustainability. Notably, they may be poor indicators of sovereign debt crises. Focusing only on the long-term debt dynamic risks overlooking the short-term factors that may influence sovereign yields and the conditions for sustainability. Early warning indictors of fiscal crises such as those developed to signal currency and banking crises should help to improve the assessment of fiscal sustainability (Kaminsky et al. 1998; Demirguc-Kunt and Detriagiache 1998). The aim of these indicators is to predict the probability of crises according to the values of a set of macroeconomic and financial variables. Although this approach seems interesting, it may still partly miss capturing the role of market sentiment, as it helps to identify the role of fundamentals and the threshold at which the crisis is triggered. In addition, computing these early warning indicators entails estimating models and consequently does not rely on a simple calculation.

Dealing with Public Debt Unsustainability: The Endogeneity Problems

Not only are interest rates assumed to be constant in fiscal sustainability indicators stemming from the IBC, but they are also assumed to be exogenous. The same hypothesis applies to the GDP growth rate. In fact, all sustainability analyses are made in a partial equilibrium setting. Although it is helpful to provide simple and easily computable indicators, it may also lead to missing critical interdependencies between macroeconomic variables. As stressed in the previous section, the empirical literature indicates that interest rates are related to market sentiment but also to fundamentals and therefore public debt. Moreover, according to the popular analyses developed by C. M. Reinhart and Rogoff (2010) and by C. M. Reinhart, V. R. Reinhart and

Rogoff (2012) based on long-term historical data, the GDP growth rate would be lowered when the public debt ratio exceeds a given threshold, estimated at 90% by C. M. Reinhart and Rogoff (2010). They therefore suggest that the growth rate is endogenously influenced by the level of public debt, but the causal relationship between public debt and growth remains a disputed issue. Finally, it should also be noticed that policy recommendations stemming from fiscal sustainability indicators have a social impact and consequences for growth, which are not taken into account in partial equilibrium analyses.

Are Interest Rates and GDP Growth Influenced by Public Debt?

The theoretical literature on fiscal policy suggests two channels through which a rise in public debt is likely to increase the interest rate. First, an expansionary policy financed by debt issuance increases the supply of public bonds and triggers portfolio adjustments. Therefore, with other things being equal, the equilibrium on the bond market is restored by an increase in bond prices and an increase in the interest rate. This is the crowding-out effect of an expansionary fiscal policy, which results in a reduction in private investment. Second, if the rise in public debt triggers an increase in the risk of sovereign default, it will amplify the rise in the interest rate as bondholders will require a risk premium to compensate for the increase in sovereign risks.

Whether or not these effects are significant is an empirical issue. The literature on the determinants of interest rates emphasises significant but weak effects of fiscal variables on interest rates. Gruber and Kamin (2012) and Poghosyan (2012) find that a one-point increase in the public debt ratio would raise interest rates by less than two basis points in a panel of OECD countries. With the sovereign debt crisis, new empirical analyses on the main drivers of sovereign spreads have been carried out. They may be particularly useful to analyse the role of sovereign risk in the determination of interest rates. A striking feature has not only been the sharp rise in spreads since 2010 but also the muted consequences of diverging debt levels among European countries before the crisis. De Grauwe and Ji (2013) confirm that fiscal variables did not have significant effects on spreads before 2008,[8] while the coefficients on debt became large and highly significant afterwards. Moreover, the impact of fiscal variables may also be non-linear, increasing with the size of debt. It must be stressed that the rise in spreads was not only the consequence of rising debt but also mirrored the risk of an exit from the euro area for countries in the periphery. The risk of sovereign default and break-up were largely

interconnected, and the financial markets may have implicitly reconsidered the exchange rate risk. Discontinuity in the pricing of sovereign default is also emphasised by Favero and Missale (2012), who show that fiscal variables only matter under special circumstances: in periods of stress when the global risk factors increase.

Consequently, the evidence on the role of fiscal variables and particularly the impact of public debt on interest rates seems to be significant but weak. This role would arise under special conditions only. The recent sovereign debt crisis episode suggests that a new regime has taken over in the euro area. Financial markets may have become more vigilant and sensitive to fiscal variables after a long period during which the sovereign risk was ignored. Recent evidence may also indicate that sovereign spreads also hinge on institutional features. As long as default and exit remain an option in the euro area, the financial markets will still price the risk of default. New institutional arrangements such as Eurobonds may be a way to avoid overshooting in crisis periods (Favero and Missale 2012) and therefore to restore the muted impact of fiscal variables on interest rates.

While the impact of public debt on the interest rate seems to be limited, C. M. Reinhart and Rogoff (2010) have recently suggested that beyond a 90% debt ratio, the side effects of public debt on growth could be significant. This finding has benefited from much attention in the recent period characterised by increased debt ratios. The Great Recession has indeed been followed by sharp increases in public debt in all industrialised countries. This increase resulted from automatic stabilisers,[9] expansionary fiscal policies and from the bailout of failing banking systems. Beyond questions about a possible risk of insolvency, the finding by C. M. Reinhart and Rogoff echoes worries about a likely decline in growth. This negative link could be associated with the crowding-out effect. If higher debt increases interest rates, it will reduce private investment and growth, not only in the short run but also in the long run if productive investment is falling. Therefore, as illustrated by equation (20.7), fiscal sustainability indicators can indicate a deterioration stemming from a higher debt ratio but also from a weaker long-term growth rate and higher interest rates. It is therefore crucial to take into account these potential feedback effects of public debt on other variables of interest. Otherwise, there would be a risk of underestimating the risk of unsustainability.

The conclusions and the results of C. M. Reinhart and Rogoff (2010) have, however, been contested. Not only, as is often the case in economics, should one not confuse correlation and causality. Weak growth may also have contributed to increasing the level of public debt. C. M. Reinhart and Rogoff (2010) and C. M. Reinhart, V. R. Reinhart and Rogoff (2012) establish a correlation but do not correct for endogeneity problems. Once this

correction is made, the correlation disappears (Panizza and Presbitero 2014). More importantly, Herndon et al. (2013) find several errors in the results published by C. M. Reinhart and Rogoff, significantly challenging the existence of a constant and stable threshold beyond which public debt would be negatively correlated with growth. Although this relation is crucial in the analysis of fiscal sustainability, the negative impact of public debt on growth remains largely a disputed issue.

What lessons can be drawn from this discussion? It should be kept in mind that the choice of the values for the interest rate and the growth rate, which are central to fiscal sustainability analysis, when computing indicators is a tricky issue, as these variables are not exogenous and may be determined by the value of debt. To provide a sound analysis of sustainability, it would be necessary to take into account the feedback effect of a debt reduction (if a policy is considered unsustainable and leads to a policy change aiming at reducing public debt) on the interest rate. Although the empirical literature on these effects has not yet reached clear conclusions, these considerations emphasise that a partial equilibrium setting misses some important issues when assessing fiscal sustainability through simple indicators.

Some Unpleasant Policy Implications of Potential Unsustainability

Beyond the possible link between the debt level and growth, the economics literature has long debated the impact of fiscal policy. As the role of fiscal sustainability indicators is to trigger a policy change—a fiscal consolidation when a policy is thought to be unsustainable—it is necessary to take into account the consequences of the change induced in fiscal policy. Fiscal consolidation may hurt growth and increase the unemployment rate. Consequently, policymakers face a trade-off between distinct and interconnected macroeconomic objectives: debt stability and employment. This implies that the analysis of fiscal sustainability cannot be realised independently of other macroeconomic objectives. Moreover, the impact of fiscal consolidation on output creates a feedback effect on the debt and public deficit dynamic through automatic stabilisers. Policy recommendations resulting from fiscal sustainability indicators exert an influence on macroeconomic variables and on public finances.

The nature of the trade-off and the importance of feedback effects critically hinge on the fiscal multiplier measuring the output effect of a one-GDP-point fiscal effort (either a one-point reduction in public spending or a one-point increase in fiscal taxes). In some extreme cases—for very high values of

the fiscal multiplier—public debt increases after fiscal consolidation, which then becomes self-defeating (Holland and Portes 2012). Although indicators point to the unsustainability of fiscal policy, they would imply solutions that would deteriorate sustainability, at least in the short term. The lessons drawn from such an indicator are therefore flawed, notably because sustainability is not being assessed in a more general theoretical framework that takes into account the interdependencies between macroeconomic variables and public finance.

The trade-off and the feedback effect of fiscal adjustment on debt are analysed by Blot et al. (2014b) through simulations of a simple macroeconomic model where the main macroeconomic interdependencies are taken into account. The model describes the dynamics of aggregate demand, potential output, the inflation rate, the monetary policy stance and public debt. Aggregate demand is driven by the fiscal effort, the interest rate and external demand. The effect of fiscal policy depends on the size of the fiscal multiplier, which is assumed to be positive and time-varying. Inflation depends on past inflation, expected inflation, the output gap and imported inflation. Monetary policy is described by a simple rule where the interest rate is supposed to increase with the output gap and with the gap between current inflation and the inflation target. As the nominal interest rate cannot be negative, the monetary policy becomes non-linear with a zero lower bound (ZLB) constraint. Finally, the equation for public debt is similar to equation (20.2), where the primary fiscal balance is decomposed as the sum of the cyclical balance and the cyclically adjusted balance (ssp_t).

The model accounts for the effect of fiscal adjustment on the output and on the debt ratio. For any positive value of the fiscal multiplier, a tax increase or a reduction in public expenditure reduces the output gap and public debt, except for extreme values of k, creating a trade-off. Moreover, the feedback effect of the adjustment is also taken into account through the role of automatic stabilisers: a decrease in the output gap following the fiscal adjustment increases the cyclical deficit and increases public debt.

The size of the multiplier is a crucial hypothesis, but this is still widely debated in the empirical literature. A review of this literature is beyond the scope of this chapter, but the simulations realised in Blot et al. (2014a, b) adopt the hypothesis of a positive value. It is also assumed that the size of the fiscal multiplier increases in times of crises and when the unemployment rate is higher. The list of factors justifying the fiscal multiplier being non-linear includes the ZLB (Eggertsson 2010), financial stress for households and firms (Corsetti et al. 2012), unemployment, the business cycle (Auerbach and Gorodnichenko 2012) and public debt thresholds (Corsetti et al. 2013).

Table 20.1 Sustainability of public debt in euro area countries

	Public debt (% of GDP)			Cumulated fiscal impulse	Output gap		GDP growth rate (%)
	2012	2020	2032	2013–2015[a]	Maximum	2013–2020	2013–2015
Germany	82	59	26	−0.3	−0.7	−0.2	1.5
France	90	82	52	−2.9	−6.8	−3.5	1.3
Italy	126.5	91	17	−2.1	−6.5	−2.4	1.3
Spain	86.1	97	83	−4.3	−9.7	−5.2	0.6
The Netherlands	69	64	49	−2.9	−2.8	−1.6	1.6
Belgium	100	80	37	−2.2	−4.3	−2.3	1.9
Portugal	119	118	82	−4.7	−10.1	−4.1	0.2
Ireland	118	133	105	−5.7	−10.9	−6.5	−0.3
Greece	177	178	93	−7.5	−17.1	−11.0	−1.5
Austria	75	62	40	−1.9	−0.9	−0.3	1.7
Finland	53	36	7	−1.3	−1.9	−0.5	2.5
Euro area	93	78	43	−2.2	−4.8	−2.4	1.0

Source: Blot et al. (2014b)

[a] Fiscal impulses are null beyond 2015

A general conclusion of this literature is that the fiscal multiplier is higher in times of crisis than in good times (see the recent meta-analysis by Gechert and Rannenberg 2014).

Simulations of the model may first be used to assess whether euro area countries are able to reach 60% debt over a 20-year horizon, given the initial values for public debt (in 2012). It can be shown that despite significant reductions in public debt ratios for all the euro area countries, fiscal policy should not be considered sustainable if the objective of 60% is retained to assess sustainability (see Table 20.1). This would be the case for Spain, Portugal, Greece and Italy. It is important to stress that such a definition is stricter than the definition proposed by Blanchard et al. (1990) according to which fiscal policy should converge to its initial value, which is the case for all the euro-area countries. However, not all the liquidity problems were solved in 2012 and Spain, Italy, Greece, Ireland and Portugal still faced difficulties in raising funds in the financial markets. The sovereign spreads with Germany were still at record levels.

With this baseline scenario, Blot et al. (2014b) investigate whether it is possible for the euro area countries to comply with the fiscal rules. To this end, they compute the additional effort needed to reach the 60% debt-to-GDP ratio over a 20-year horizon. The additional effort is a set of −0.5 or +0.5 fiscal impulses over 2015–2032 until each member state reaches the 60% target. Thus, they aim to gauge whether all the countries can reach the public debt target in 2032. For countries that have already achieved this threshold, they

Table 20.2 Is it possible to get to 60% in 2032?

	Public debt in 2032	Average GDP growth (2013–2020)[a]	Cumulated fiscal impulse
Germany	58	0.2	2.2
France	60	0.2	−1.9
Italy	58	0.2	1.9
Spain	55	−0.6	−9.3
Netherlands	55	0.0	−2.4
Belgium	55	0.1	−0.7
Greece	85	−1.2	−14.2
Portugal	57	−0.5	−7.7
Ireland	71	−0.4	−16.0
Austria	55	0.0	−0.9
Finland	58	0.1	3.2
Euro area	58	0.0	−1.2

Source: Blot et al. (2014b)
[a]As the difference from the central scenario summarised in Table 20.1

implement positive fiscal impulses after 2015 so that debt-to-GDP ratio is equal to 60% in 2032. The positive or negative 0.5 percentage point ticks are fixed for the sake of simplicity. The fiscal impulse is negative (positive) if the actual debt is above (below) the target.

Spain and Portugal would be able to reach the 60% objective for the debt ratio if they implement further consolidation after 2015 (Table 20.2). The additional fiscal effort would reach five points of GDP for Spain and three points for Portugal. These figures may be interpreted as the tax gap (or expenditure gap) in equation (20.8). However, the model also accounts for the output costs of the fiscal adjustment, and it is shown that the annual average growth rate would decline by 0.6 and 0.5 in Spain and Portugal, respectively. Taking into account macroeconomic interdependencies is then essential in order to illustrate trade-offs and a potential conflict of objectives. Satisfying the debt criteria would be a costly strategy and would impede countries from recovering from the crisis, since unemployment would remain above its equilibrium value (Blot et al. 2014b).[10] It is also shown that complying with fiscal rules would not be feasible for Greece and Ireland. Despite a significant additional fiscal effort, the debt ratio would still be above 60%, and average growth would be lowered from 2015 up to 2032. Beyond these short-term negative impacts on growth, cutting investment expenditure may also lower future growth.[11]

Fiscal adjustment under these conditions may appear unrealistic, given the historical level of primary fiscal surpluses, and unreasonable (see International Monetary Fund 2012). Between 2013 and 2017, Greece, Spain and Portugal would experience slower economic growth entailing dramatic social costs.

Even for other countries, reaching a 60% debt-to-GDP ratio would imply keeping the primary fiscal balance at high historical levels. In fact, it is not sure that maintaining so high a level of fiscal adjustment is feasible. Buiter and Rahbari (2013) suggest considering the maximum value of the primary surplus that can be reached taking into account economic, social and political factors. Beyond its social costs, austerity can become politically unsustainable. It may also have dramatic consequences for health, as suggested by several medical studies (see Stuckler and Basu 2013).[12] The IBC can help to determine the minimum primary surplus required to stabilise debt and can be compared to the maximum value to assess whether an adjustment is feasible or not. Although there is no metric for this maximum value, this argument clearly shows that a consistent view of sustainability should also encompass other objectives. This is the only way for policymakers to address the conflicts between social and economic objectives. Fiscal sustainability indicators are, therefore, useful, but assessments of fiscal sustainability should be complemented with additional indicators reflecting other policymakers' objectives in order to avoid the long-term objectives (public debt sustainability) being favoured at the expense of short-term objectives.

Conclusion

The theoretical background for sustainability indicators stems from governments' IBC. Contrary to hypothesis testing of the validity of the IBC or the transversality condition, indicators aim to provide a simple and clear message about the situation of public finances. Although they are based on a theory of public debt sustainability, the reference value used in fiscal sustainability indicators lacks a solid basis. There is no reason to suppose that the public debt ratio should converge to an ad-hoc exogenous value in the future (the initial value or any other reference value). According to the IBC, governments may be solvent for any value of debt as long as the current level of debt is backed by future primary surpluses.

The main criticisms of these indicators are related to the fact that they are derived from a partial equilibrium setting. Consequently, fiscal sustainability indicators do not account for macroeconomic interdependencies. In particular, they assume that the interest rate and the GDP growth rate are constant and exogenous. The main message of this chapter is that this view risks overlooking the important feedback effect of public debt and of the policy response—fiscal adjustment—which is implemented when fiscal policy is shown to be unsustainable. In particular, the risk of public default may be

underestimated, as it sometimes results from liquidity problems, which are not considered in existing indicators. It is important to bear in mind that a government is solvent as long as other agents are ready to acquire new debt issued by the government.

Furthermore, addressing sustainability may prevent policy-makers from achieving other objectives. Restoring public debt sustainability implies consolidating public finance, which entails output losses. There is therefore a need to take into account macroeconomic interdependencies to address a potential conflict of objectives. This can only be done with a more general equilibrium perspective. Overlooking these trade-offs risks undesirable economic and social outcomes as has been seen with the recent episode of fiscal consolidation in the euro area. Fiscal adjustments were implemented to reinforce sustainability but with huge output and social costs.

Notes

1. Primary expenditure is equal to total current expenditure less interest payments on debt.
2. A Ponzi game is a situation where the debtor issues new debt to pay interests on past debt. It comes from the fraudulent investment operation proposed by Charles Ponzi in 1920, who proposed a high return on investment but was only able to pay early investors with funds received from new investors.
3. The output gap is the difference between the actual output and the potential output. The potential output for an economy is not yet observed and can only be estimated.
4. These measures are computed by the main European or international institutions (the OECD, the IMF and the European Commission). However, it must be stressed that these evaluations are still subject to uncertainty regarding the exact measure of the output gap, as the potential level of activity (or growth) is not observed.
5. The indicator may also serve to account for the impact of reforms of the healthcare or pension systems.
6. The idea that we need a more relevant model to assess sustainability of fiscal policy developed by Bohn's theoretical approach remains generally unsatisfactory.
7. See Art. 126 of the Treaty and Protocol (N°12).
8. Spreads were higher before European Economic and Monetary Union (EMU) but resulted mainly from exchange rate risks.
9. Deficit rises automatically when the output gap is negative as weakening growth reduces fiscal revenues while rising unemployment increases public spending and notably social expenditures (mainly unemployment benefits).

10. Blot et al. (2014a) suggest that alternative strategies should be contemplated. Taking into consideration the time-varying properties of the fiscal multiplier, it would be optimal to delay the fiscal adjustment. The output costs would be minimised, while the debt ratio may still reach 60%.
11. This impact is not taken into account in the simulations presented in Blot et al. (2014a, b).
12. Stuckler and Basu (2013)'s research on health and social policy, which analyses the consequences of fiscal consolidation experiences during the Great Depression and the post-communist period in Iceland and Greece.

References

Aldama, P. & Creel, J. (2016). *Why Fiscal Regimes Matter for Fiscal Sustainability Analysis: An Application to France* (Observatoire Français des Conjunctures Économiques, OFCE Working Paper N°15). Paris: Sciences Po.

Altavilla, C., Giannone, D. & Lenza, M. (2014). *The Financial and Macroeconomic Effects of OMT Announcements* (European Central Bank Working Paper Series 1707). Frankfurt am Main: European Central Bank.

Arghyrou, M. G., & Kontonikas, A. (2012). The EMU Sovereign-Debt Crisis: Fundamentals, Expectations and Contagion. *Journal of International Financial Markets, Institutions and Money, 22*(4), 658–677.

Arghyrou, M. G., & Tsoukalas, J. D. (2011). The Greek Debt Crisis: Likely Causes, Mechanics and Outcomes. *World Economy, 34*, 173–191.

Auerbach, A. J., & Gorodnichenko, Y. (2012). Measuring the Output Responses to Fiscal Policy. *American Economic Journal: Economic Policy, 4*(2), 1–27.

Balassone, F., & Franco, D. (2000). *Assessing Fiscal Sustainability: A Review of Methods with a View to EMU (Banca D'Italia Fiscal Sustainability)*. Rome: Banca D'Italia, Servizio Studi. Retrieved from: https://www.bancaditalia.it/pubblicazioni/altri-atti-convegni/2000-fiscal-sustainability/021-060_balassone_and_franco.pdf?language_id=1.

Balassone, F., Cunha, J., Langenus, G., Manzke, B., Pavot, J., Prammer, D., & Tommasino, P., (2009). *Fiscal Sustainability and Policy Implications for the Euro Area* (European Central Bank Working Paper Series, No. 994). Frankfurt am Main: European Central Bank. Retrieved from: https://www.ecb.europa.eu/pub/pdf/scpwps/ecbwp994.pdf?1592c06cb2ccae7468e749a7e2ec9b87

Bernoth, K., Von Hagen, J., & Schuknecht, L. (2012). Sovereign Risk Premiums in the European Government Bond Market. *Journal of International Money and Finance, 31*(5), 975–995.

Berrittella, M., & Zhang, J. (2015). Fiscal Sustainability in the EU: From the Short-Term Risk to the Long-Term Challenge. *Journal of Policy Modelling, 37*, 261–280.

Blanchard, O. (1990). *Suggestions for a New Set of Fiscal Indicators* (Organisation for Economic Co-operation and Development [OECD] Working Paper No. 79).

Paris: OECD. Retrieved from: http://www.oecd-ilibrary.org/economics/suggestions-for-a-new-set-of-fiscal-indicators_435618162862

Blanchard, O., Chouraqui, J. C., Hagemann, R. P., & Sartor, N. (1990). *The Sustainability of Fiscal Policy: New Answers to an Old Question* (OECD Economic Studies No. 15). Paris: OECD. http://www.oecd-ilibrary.org/economics/suggestions-for-a-new-set-of-fiscal-indicators_435618162862

Blot, C., Cochard, M., Creel, J., Ducoudré, B., Schweisguth, S., & Timbeau, X. (2014a). Fiscal Consolidation in Times of Crisis: Is the Sooner Really Better? *Revue de l'OFCE, 132*, 159–192.

Blot, C., Cochard, M., Creel, J., Ducoudré, B., Schweisguth, D., & Timbeau, X. (2014b). Fiscal Consolidation, Public Debt and Output Dynamics in the Euro Area: Lessons from a Simple Model with Time-Varying Fiscal Multipliers. *Revue d'Economie Politique, 124*(6), 953–989.

Blot, C., Ducoudré, B., & Timbeau, X. (2016). Sovereign Spread and Default in a Model with Self-Fulfilling Prophecies and Asymmetric Information. *Journal of Macroeconomics, 47*, 281–299.

Bohn, H. (1995). The Sustainability of Budget Deficits in a Stochastic Economy. *Journal of Money, Credit and Banking, 27*(1), 257–271.

Bohn, H. (1998). The Behaviour of U.S. Public Debt and Deficits. *The Quarterly Journal of Economics, 113*(3), 949–963.

Bohn, H. (2007). Are Stationarity and Cointegration Restrictions Really Necessary for the Intertemporal Budget Constraint? *Journal of Monetary Economics, 54*, 1837–1847.

Bohn, H. (2008). The Sustainability of Fiscal Policy in the United States. In R. Neck & J. E. Sturm (Eds.), *Sustainability of Public Debt* (pp. 15–49). Cambridge, MA: MIT Press.

Buiter, W., & Rahbari, E. (2013). *Why Do Governments Default and Why Don't They Default More Often?* (Center for Economic Policy Research [CEPR] Discussion Paper No. 9492). London: CEPR. Retrieved from: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2271943

Calvo, G. (1988). Servicing the Public Debt: the Role of Expectations. *American Economic Review, 78*(4), 647–661.

Chalk, N. & Hemming, R. (2000). *Assessing Fiscal Sustainability in Theory and Practice* (International Monetary Fund [IMF] Working paper 00/81). Washington, DC: IMF.

Cole, H. L., & Kehoe, T. J. (1996). A Self-Fulfilling Model of Mexico's 1994–1995 Debt Crisis. *Journal of International Economics, 41*, 309–330.

Corsetti, G., Meier, A., & Müller, G. J. (2012). What Determines Government Spending Multipliers? *Economic Policy, 27*(72), 521–565.

Corsetti, G., Kuester, K., Meier, A., & Müller, G. J. (2013). Sovereign Risk, Fiscal Policy and Macroeconomic Stability. *Economic Journal, 123*, F99–F132.

Davig, T. (2005). Periodically Expanding Discounted Debt: A Threat to Fiscal Policy Sustainability? *Journal of Applied Econometrics, 20*(7), 829–840.

De Grauwe, P. (2012). The Governance of a Fragile Eurozone. *Australian Economic Review, 45*(3), 255–268.

De Grauwe, P., & Ji, Y. (2013). Self-Fulfilling Crises in the Eurozone: An Empirical Test. *Journal of International Money and Finance, 34*, 15–36.

De Grauwe, P., & Ji, Y. (2014). How Much Fiscal Discipline in a Monetary Union? *Journal of Macroeconomics, 39*, 348–360.

Demirguc-Kunt, A., & Detragiache, E. (1998). *The Determinants of Banking Crises in Developing and Developed Countries* (IMF Staff Papers, 45/1, 81–109). Washington, DC: IMF.

Diamond, D., & Dybvig, P. (1983). Bank Runs, Deposit Insurance, and Liquidity. *Journal of Political Economy, 91*(3), 401–419.

Doi, T., Hoshi, T., & Okimoto, T. (2011). Japanese Government Debt and Sustainability of Fiscal Policy. *Journal of the Japanese and International Economies, 25*, 414–433.

Eggertsson, G. (2010). What Fiscal Policy is Effective at Zero interest Rates? *National Bureau of Economic Research Macroeconomics Annual, 25*, 59–112.

European Central Bank. (2012). Analysing Government Debt Sustainability in the Euro Area. *ECB Monthly Bulletin*, April, 55–69.

European Commission. (2015). *The 2015 Ageing Report. Economic and Budgetary Projections for the EU28 Member States (2013–2060)*. European Economy, 3. Retrieved from: http://ec.europa.eu/economy_finance/publications/european_economy/2015/pdf/ee3_en.pdf

Favero, C., & Missale, A. (2012). Sovereign Spreads in the Eurozone: Which Prospects for a Eurobond? *Economic Policy, 27*(70), 231–273.

Gechert, S. & A. Rannenberg, A. (2014). *Are Fiscal Multipliers Regime-Dependent? A Meta-Regression Analysis* (Institut für Macroökonomie und Konjunkturforschung Working Paper 139). Retrieved from: https://www.boeckler.de/pdf/p_imk_wp_139_2014.pdf

Ghosh, A. R., Kim, J. I., Mendoza, E. G., Ostry, J. D., & Qureshi, M. S. (2013). Fiscal Fatigue, Fiscal Space and Debt Sustainability in Advanced Economies. *The Economic Journal, 123*(566), F4–F30.

Giammarioli, N., Nickel, C., Rother, P., & Vidal, J. P. (2007). *Assessing Fiscal Soundness: Theory and Practice* (European Central Bank [ECB] Occasional Paper No. 56). Frankfurt am Main: ECB. Retrieved from: https://www.ecb.europa.eu/pub/pdf/scpops/ecbocp56.pdf?59fd0bc88cf8ba88e7f7846038ac96b5

Gruber, J., & Kamin, S. (2012). Fiscal Positions and Government Bond Yields in OECD Countries. *Journal of Money, Credit and Banking, 44*(8), 1563–1587.

Herndon, T., Ash, M., & Pollin, R. (2013). Does High Public Debt Consistently Stifle Economic Growth? A Critique of Reinhart and Rogoff. *Cambridge Journal of Economics, 38*(2), 257–270.

Holland, D., & Portes, J. (2012). Self-Defeating Austerity? *National Institute Economic Review, 222*(1), F4.

IMF. (2012, October). The Good, the Bad and the Ugly: 100 Years of Experience of Dealing with Debt Overhang. In IMF (Ed.), *IMF Economic Outlook* (pp. 101–127). Washington, DC: IMF.

Kaminsky, G., Lizondo, S., & Reinhart, C. M. (1998). *Leading Indicators of Currency Crises* (IMF Staff Papers, 45/1, 1–48). Washington, DC: IMF. Retrieved from: https://www.imf.org/external/pubs/ft/staffp/1998/03-98/pdf/kaminsky.pdf

Laeven, L., & Valencia, F. (2012). *Systemic Banking Crises Database: An Update* (IMF Working Paper 12/163). Washington, DC: IMF. Retrieved from: https://www.imf.org/external/pubs/cat/longres.aspx?sk=26015.0

McHale, J. (2012). An Overview of Developments in the Irish Economy Over the Last Ten Years. *The World Economy, 35*(10), 1220–1238.

Oosterlinck, K. (2013). Sovereign Debt Defaults: Insight from History. *Oxford Review of Economic Policy, 29*(4), 697–714.

Panizza, U., & Presbitero, A. (2014). Public Debt and Economic Growth: Is There a Causal Effect? *Journal of Macroeconomics, 41*, 21–41.

Poghosyan, T. (2012). *Long-Run and Short-Run Determinants of Sovereign Yields in Advanced Economies* (IMF Working Paper No. 12/271). Washington, DC: IMF. Retrieved from: https://www.imf.org/external/pubs/cat/longres.aspx?sk=40091.0

Reinhart, C. M., & Rogoff, K. S. (2010). Growth in a Time of Debt. *American Economic Review, 100*(2), 573–578.

Reinhart, C. M., Reinhart, V. R., & Rogoff, K. S. (2012). Public Debt Overhangs: Advanced-Economy Episodes Since 1800. *Journal of Economic Perspectives, 26*(3), 69–86.

Stuckler, D., & Basu, S. (2013). *The Body Economic: Why Austerity Kills*. London: Allen Lane.

Whelan, K. (2014). Ireland's Economic Crisis: The Good, the Bad and the Ugly. *Journal of Macroeconomics, 39*, 424–440.

21

Measuring the World: Indicators, Human Rights, and Global Governance

Sally Engle Merry

Introduction

Indicators are rapidly multiplying as tools for assessing and promoting a variety of social justice and reform strategies around the world. There are rule of law indicators, indicators of violence against women, and indicators of economic development, among many others. Indicators are widely used at the national level and are increasingly important in global governance. Although the origins of indicators as modes of knowledge and governance stretch back to the creation of modern nation states in the early nineteenth century and practices of business management a few centuries earlier, their current use in global governance comes largely from economics and business management. Development agencies such as the World Bank have created a wide range of indicators, including indicators of global governance and the rule of law, while gross domestic product is one of the most widely used and accepted

First published in Partridge, D. J., Welker, M., & Hardin, R., (2011). *Current Anthropology*, 52 (3), (Supplement to April 2011). *Corporate Lives: New Perspectives on the Social Life of the Corporate Form*, (pp. S83–S95), this chapter benefited greatly from comments by the participants at the Wenner Gren/ School of Advanced Research conference on "Corporate Lives: New Perspectives on the Social Life of the Corporate Form". I am grateful to Joseph Weiler and the student participants of International Institute of Law and Justice Colloquium at the New York University Law School, and to Richard Abel, Jane Anderson, and Frank Munger for comments on an earlier draft. This research has been generously supported by the National Science Foundation Law and Social Sciences Program and Science, Technology and Society Program. #SES-0921368

S.E. Merry (✉)
New York University, New York, NY, USA

D.V. Malito et al. (eds.), *The Palgrave Handbook of Indicators in Global Governance*,
https://doi.org/10.1007/978-3-319-62707-6_21

indicators. Thus, the growing reliance on indicators is an instance of the dissemination of the corporate form of thinking and governance into broader social spheres. They are fundamental to modern forms of governmentality, whether in the service of corporate, state, or reform modes of governance.

Although indicators are widely used in reform initiatives at the global level under the auspices of the United Nations (UN) and international NGOs, they are also increasingly important to corporate social responsibility initiatives. The UN Global Compact (GC) and the Global Reporting Initiative (GRI) are two of the most significant entities promoting corporate social responsibility, and both rely increasingly on indicators. There are also NGOs developing toolkits to measure corporate compliance with human rights standards. However, in accordance with contemporary audit culture (see Power 1999), these efforts place responsibility for gathering information and assessing it on the organizations themselves. The GRI also provides for stakeholder discussions of the relevant indicators that they will use, and some of the human rights toolkits are flexible. Clearly, this approach to monitoring faces problems of verifying the information it uses, given its reliance on self-reporting and even on choice of measures. These systems are all voluntary, monitored only by a corporation's concern for its public respectability, and reputation.

One of the fascinating revelations of the Corporate Lives seminar was recognizing how much corporations participate in the same forms of identity formation as individuals. Corporations work to construct desirable reputations, investing substantial resources in maintaining them through advertising and self-monitoring. Consumer movements have ratcheted up the financial consequences of corporate social irresponsibility by boycotting goods produced by irresponsible corporations and labeling or certifying responsible corporations. An irresponsible corporate performer tarnishes all the other corporations in the same field. It is ironic that the power of indicators, and their monitoring and accountability mechanisms, is primarily dependent on their impact on corporate reputations. However, as the other chapters in this collection indicate, the corporate form shapes the way individuals are understood in the current period, so it is not surprising that corporations are reciprocally understood as social beings with identities and reputations.

The Expansion of Indicators for Global Governance

Technologies of audit and performance evaluation common in the corporate world now reach into many domains of global governance. Since the mid-1990s, technologies that were developed in the sphere of business regulation

have jumped domains to human rights and corporate social responsibility. Interest in using indicators to monitor human rights compliance has grown significantly. Indicators introduce into the field of global human rights law a form of knowledge production in which numerical measures make visible forms of violation and inequality that are otherwise obscured. Statistics on income, health, education, and torture, for example, are useful to assess compliance with human rights norms and progress in improving human rights conditions. The use of these statistics, and indicators derived from them, by the committees charged with monitoring compliance with the major human rights conventions has increased over the past two decades. Some committees, as well as the UN's Office of the High Commissioner of Human Rights, are developing more sophisticated indicators to facilitate the analysis of information and increase accountability. Indicators, particularly those that rely on ranks or numbers, convey an aura of objective truth and facilitate comparisons. However, indicators typically conceal their political and theoretical origins and underlying theories of social change and activism. They rely on practices of measurement and counting that are themselves opaque.

The world of civil society organizations has also been transformed by the increasing use of statistical measures. There are demands for quantifying the accomplishments of civil society organizations and for "evidence-based" funding. Donors to human rights organizations want indicators of success such as reductions in trafficking in persons, or diminished rates of poverty and disease. As donors move closer to business, they have adopted business-based means of accounting for productivity and accomplishments. The concept of "venture philanthropy" underscores this new perspective. Recipient organizations are tasked to develop measures of what they have accomplished within the period of funding. Given the difficulties of measuring accomplishments such as "increased awareness of human right", NGOs tend to count proxies for these accomplishments, such as number of training sessions or number of people trained. Clearly, the use of quantitative measures of accomplishment and the introduction of ranking systems based on these measures is transforming the way these organizations do their work.

This chapter considers two sociological aspects to the expansion of the use of indicators. The first is a knowledge effect. Numerical measures produce a world knowable without the detailed particulars of context and history. The constituent units can be compared and ranked according to some criteria. This knowledge is presented as objective and often as scientific. The interpretations lurk behind the numbers but are rarely presented explicitly. These numbers seem open to public scrutiny and readily accessible in a way that private opinions are not. The second is a governance effect. Statistical measures of populations are clearly connected to eighteenth- and early

nineteenth-century ideas that the people of a country represent its wealth, and that good governance requires measuring and counting these people.

As forms of knowledge, indicators rely on the magic of numbers as well as the appearance of certainty and objectivity that they convey. A key dimension of the power of indicators is their capacity to convert complicated, contextually variable phenomena into unambiguous, clear, and impersonal measures. They represent a technology of producing readily accessible and standardized forms of knowledge. Indicators are a special use of statistics to develop quantifiable ways of assessing and comparing characteristics among groups, organizations, or nations. They depend on the construction of categories of measurement such as ethnicity, gender, income, and more elaborated concepts such as national income. Indicators submerge local particularities and idiosyncrasies into universal categories, generating knowledge that is standardized and comparable across nations and regions.

One of the critical ways an indicator produces knowledge is by announcing what it measures, such as "rule of law" or "poverty". Neither of these categories is self-evident. When sponsoring organizations name their indicators, they interpret what the numbers mean. Labeling is essential to produce a measure that is readily understood by the public and simple in its conception. Labels do not necessarily accurately reflect the data that produce the indicators, however. How indicators are named and who decides what they represent are fundamental to the way an indicator produces knowledge.

Indeed, statistical measures create new categories. An indicator may even create the phenomenon it is measuring instead of the other way around. IQ is whatever it is that the IQ test measures, for example. Here, the process of measurement produces the phenomenon it claims to measure. As Porter points out, although the categories of enumeration may be highly contingent at first, once they are in place, they become extremely resilient and come to take on permanent existence as a form of knowledge. He uses the category of Hispanic in the US census as an example of this phenomenon (Porter 1995, p. 42). One of the most well-known examples of this process is the introduction of the census in India by the British colonial authorities in the nineteenth and twentieth centuries (Cohn 1996; Randeria 2006; Dirks 2001). To increase legibility, the population census classified individuals by caste, religion, gender, and other criteria. The British arranged the castes in an orderly hierarchy and sought to collect "objective" information about caste identities. However, the caste categories in existence at the time were relatively fluid, situational, segmented, and local. In place of a wide range of forms of ritual and social exclusion in practice, the British selected pollution by touch as the key marker of low caste status. Thus, the category "Untouchabilty" emerged as a distinct,

all-India category. By redefining castes in terms of categories that applied across the subcontinent, the British rendered caste into a far more fixed and intractable social entity, but one that could be more readily counted and compared (Randeria 2006, p. 19).

Indicators are not only a technology of knowledge production but also of governance. They are widely used for decisions such as where to send foreign aid, where to focus on human rights violators, and which countries offer the best conditions for business development. Modern states use statistical information, some of which is bundled into indicators, to decide where to locate highways and railroads, where to build schools and hospitals, how to allocate taxes, and how to deploy police forces to control crime, to give only a few examples. As the modern state came to see its wealth as its population, it put greater emphasis on counting and assessing the nature of the population. Standardized measures mean the state can better administer its population, by knowing its birth and death rates as well as income levels, for example, and collect taxes (Porter 1995, p. 25).

The use of statistical information, in general, and indicators in particular, shifts the power dynamics of decision-making. Indicators replace judgments on the basis of values or politics with apparently more rational decision-making on the basis of statistical information. In theory, the process is more open, allowing the public access to the basis for decisions. As Porter argues (1995), in the pre-modern world, aristocratic elites relied on non-numerical information circulated within small, private circles. Statistical knowledge grew in importance with the birth of the modern state. The first great enthusiasm for statistics in Europe came in the 1820s and 1830s, and by the mid-nineteenth century, in France, statistics were thought to produce the broad public knowledge necessary for a democracy. Quantification provided an openness to public scrutiny. For French bridge and canal engineers at mid-century, for example, calculating public utility by numbers offered a defense against parochialism and local interests in the locations of railroads and canals (Porter 1995, p. 121). The massive expansion of quantification in recent times comes from a political culture that demands more openness and seeks to drive out corruption, prejudice, and the arbitrary power of elites even at the cost of subtlety and depth (Porter 1995, pp. 85–6). This, Porter claims, is the power of numbers.

Yet, statistical measures have embedded theories and values which shape apparently objective information and influence decisions. Despite the increase in democratic openness produced by the use of statistics in decision-making, this is a technology that tends to consolidate power in the hands of those with expert knowledge. In many situations, the turn to indicators as modes of

governance does not eliminate the role of private knowledge and elite power in decision-making, but replaces it with technical, statistical expertise. Decisions that were carried out by political or judicial leaders are made by technical experts who construct the measures and develop the processes of classification and counting that produce the numbers. In nineteenth-century France, for example, despite claims to rigorous definition and lack of ambiguity, statistical measures were often arcane and hard to understand, requiring careful interpretation by experts (Porter 1995, pp. 74, 80–1). In the area of contemporary global governance, an increasing reliance on indicators tends to locate decision-making in the global North, where indicators are typically designed and labeled.

Indicators provide a technology for reform as well as control. Indicators can effectively highlight deficits, areas of inequality, spheres of human rights violations, and other problem areas. Reform movements depend on producing statistical measures of the wrongs they hope to redress, such as human rights violations, refugee populations, disease rates, and the incidence of poverty and inequality. They are a valuable reform tool in their ability to show areas of state failure.

As indicators become increasingly central to global reform and global governance, it is critical to examine how they are produced and how the forms of knowledge they create affect global power relationships. They influence the allocation of resources, the nature of political decisions, and the assessment of which countries have bad human rights conditions. They facilitate governance by self-management rather than command. Individuals and countries are made responsible for their own behavior as they seek to comply with the measures of performance articulated in an indicator.

This chapter advocates an ethnographic approach to understanding the role and impact of indicators. Doing an ethnography of indicators means examining the history of the creation of an indicator and its underlying theory, observing expert group meetings and international discussions, where the terms of the indicator are debated and defined, interviewing expert statisticians and other experts about the meaning and process of producing indicators, observing data collection processes, and finally examining the ways indicators affect decision-making and public perceptions. I am in the early stages of an ethnographic study of three human rights indicators, tracing the social networks and systems of meaning through which they are produced and used. A critical dimension of the ethnography of global indicators is an analysis of the sources of information they use and forms of cooperation and resistance by countries and NGOs in the contest over who counts and what information counts.

Defining Indicators

Indicators are statistical measures that are used to summarize complex data into a simple number or rank that is meaningful to policymakers and the public. They tend to ignore individual specificity and context in favor of superficial but standardized knowledge. An indicator presents clearly the most important features relevant to informed decision-making about one issue or question (UN Expert Group Meeting 2007, p. 4).[1] Although indicators are quantitative, expressed in rates, ratios, or percentages, or numbers, some are based on qualitative information converted into numbers. A recent effort to develop indicators for Convention on the Elimination of all Forms of Discrimination Against Women (CEDAW), for example, uses quantitative indicators such as literacy rates, maternal mortality rates, and labor force participation rates that are sex-disaggregated, along with qualitative indicators such as the existence of legislation such as equal inheritance rights, polices such as quotas for girl children in educational institutions, and programs such as legal aid services and shelters for women victims of violence. These qualitative measures are quantified by counting the number of laws, the number of shelters, and so on, to produce a number (Goonesekere 2004, pp. 10–11). Some indicators use a variety of qualitative measures to construct an ordinal numerical ranking, as is the case with rule of law measures which assess a country's rule of law on a scale of 1–5 (Davis 2004, p. 152). Many indicators are composites of other indicators, a blending and weighting of established indicators into a new bundle (see Kaufmann and Kraay 2007).

The importance of understanding indicators emerged during my conversations about human rights reform with several senior UN staff members. They argued that it was impossible to engage in reform projects without indicators and were working to develop indicators of early marriage. They confronted conceptual challenges in determining the age of marriage. Did marriage begin at the age of betrothal, the age at the wedding ceremony, the age of first sex, or the age of cohabitation? These events have different implications for human rights violations. Age of betrothal might flag forced marriage, since younger girls are less likely to exercise free choice. Not all societies have recognizable wedding ceremonies, nor do they necessarily lead to first sex or cohabitation. Age of first sex could indicate medical complications of early childbearing such as fistula. Cohabitation might spell the end of a girl's schooling. One UN staffer sighed and noted that marriage is very complicated. Despite these complexities, they settled on cohabitation. I have since pondered this choice, thinking about the difference it would have made were another criterion chosen and wondering how the decision was made and by whom. What were

the criteria? Was it the availability of data? To what extent was this decision based on a theory of early marriage and particular health or social problems?

Indicators typically do not come with a discussion of such decisions or an analysis of the implications of the choice. Clearly, the selection of any criterion depends on how marriage is defined. Depending on which criterion is chosen, the indicator could measure how much early marriage and childbearing damage health, diminish women's schooling, or prevent free choice of partners. The indicator submerges these issues and their surrounding theories. The essence of an indicator is that it is simple and easy to understand. Embedded theories, decisions about measures, and interpretations of the data are replaced by the certainty and lack of ambiguity of a number. Like money, it appears to allow abstraction and easy comparison among groups and countries by converting values into numbers. But what information is lost? Does the number bury the messiness of difference and allow equivalence (Kaufmann and Kraay 2007)?[2]

A comparison with money is instructive, since it is the quintessential unit that flattens difference into commensurate values. The "cash nexus" famously pointed to money's capacity to make possible comparison and exchange of items, such as potatoes and sex (Poovey 1998).[3] But does money bury the messiness of difference and allow equivalence? As Bill Maurer notes, the apparent equivalence created by money is undermined by questions of morality and sociality. He examines alternative currencies, such as Islamic banking or community currency, in upstate New York that are grounded in critiques of capitalism. Although the money in each system is technically fungible with the others, translation is not simple. The currencies co-exist as convertible but socially incommensurate in meaning and morality. Efforts to move between currencies or to do Islamic banking lead to awkward compromises. Maurer refers to the operation of the uncanny as a way to think about the tension of things that are the same but always different (Maurer 2005, pp. 104–121). Indicators rely on a similar alchemy: they create a commensurability that is widely used to compare, to rank, and to make decisions, even though the users recognize that these simplified numerical forms are superficial, often misleading, and very possibly wrong.

Human Rights and Audit Culture

The use of indicators to monitor compliance with human rights is a rapidly growing field. Until the late 1990s, many human rights activists resisted the use of indicators because of concerns about lack of data, oversimplification,

and bias (see Green 2001, pp. 1082–84; Alston 2005, p. 22; Rosga and Satterthwaite 2009). For example, the Freedom House indicator, "Freedom in the World", with its seven-point scale from "free" to "not free", based on annual surveys starting in 1972, was widely seen as ideologically biased (Alston 2005, p. 23). Efforts to develop indicators for social and economic human rights have faced difficulties in making the measures concrete (Rosga and Satterthwaite 2009). Indicators measure aggregates, while human rights are held by individuals (see Green 2001, p. 1085). Building a composite index of human rights performance promotes quick comparisons of countries along a scale but ignores the specificity of various human rights and conceals particular violations. Measurement errors are also a major concern. There are significant differences in the quality of data on human rights violations among countries. Those countries more concerned about human rights are likely to report a higher proportion of violations than those that resist human rights principles (Alston 2005, pp. 22–25).

Despite these concerns, the use of indicators is growing in the human rights field, migrating from economics through development to human rights compliance. UN agencies such as UNICEF, UNIFEM, the Commission on the Status of Women, the High Commissioner on Human Rights, and the UN Statistical Commission are taking the lead. There are long-standing initiatives to develop statistical indicators among other UN agencies and programs, such as FAO, ILO, UNESCO, UNICEF, WHO, and UNDP (Malhotra and Fasel 2005). A set of indicators has been developed for the Millennium Development Goals (MDGs; Millennium Development Goals Indicators 2017). Universities and non-governmental organizations are also active in collecting and systematizing data. For example, the University of Maryland has a research project on *Minorities at Risk* that examines the status and conflicts of politically active groups (Center for International Development and Conflict Management 2017; Malhotra and Fasel 2005, p. 21). Many economic and social indicators, such as the World Bank Worldwide Governance Indicators and the UNDP Human Development Index, are used to assess compliance with social and economic human rights (Filmer-Wilson 2005, p. 28; Green 2001).

Development agencies have long used indicators. The recent shift to a rights-based approach to development (Sen 1999) has brought human rights and development closer together and encouraged the use of economically based indicators for human rights compliance. The 2000 UNDP Human Development Report devoted a chapter to the value of indicators for human rights accountability (UNDP 2000, pp. 89–111). The World Bank has collected and disseminated a wide range of socio-economic statistics derived largely from national statistical systems as well as data on governance, and the

rule of law based on expert and household surveys (Malhotra and Fasel, 2005, p. 15). These are useful for monitoring compliance with social and economic rights, in particular (Green 2001). Economists at the World Bank have also played a critical role in developing indicators for international investment, such as its *Doing Business* project to assess business conditions around the world (Davis and Kruse 2007, p. 1097). The 2009 *Doing Business Report* ranked 181 countries on ten criteria for doing business, such as starting a business or dealing with construction permits, producing an overall Ease of Doing Business Index (2016). Singapore ranked first, the USA third, and the Democratic Republic of the Congo last. The website offers a one-page explanation of the index and a caution about its limited scope. Despite these limitations, the index offers a readily understandable comparative exposition of business conditions around the world in one short table.

In his anthropological account of a European development project in Africa, Richard Rottenburg uses Latour's concept of centers of calculation to describe the production of such comparative, translocal knowledge. In order for a development bank to produce the knowledge necessary to monitor and control projects, it must know about projects around the world in comparable terms through their reports. Bank officials juxtapose these reports to create a common context that produces new knowledge. The process depends on producing representations of projects that travel (reports), that are immutable (certain in meaning, not shifting according to the teller), and combinable. Making reports combinable requires establishing in advance standardized procedures for measuring and aggregating the information in the report. By comparing the reports, the development bank produces translocal knowledge that allows it to monitor and control projects from a distance and to be accountable to the taxpayers (Latour 1987; Rottenburg 2009, pp. 181–2). This process, developed in the domain of economics and reliant on universalistic technical standards, provides a template for the production and use of indicators in other domains.

While there is considerable discussion of how to develop good indicators and critiques of their errors of measurement, quality of data, embedded assumptions, and simplification (see Davis 2004), there is far less attention to the implications of the use of indicators for practices of global governance itself (but see Rosga and Satterthwaite 2009). Within social science, however, there has been considerable attention to the impact on practices of governance of these new political technologies based on statistics and accountability, what has been called "audit culture" (Power 1999; Strathern 2000). Audit technologies are theorized as instruments for new forms of governance and power, "agents for the creation of new forms of subjectivity: self-managing

individuals who render themselves auditable" (Shore and Wright 2000, p. 57). These technologies allow people to check their behavior for themselves so that governments can withdraw from checking behavior and simply check indicators of performance (Strathern 2000, p. 4). The self-checking practices become evidence of accountability from the perspective of the state. Marilyn Strathern's edited collection focuses on new mechanisms for accountability established by the British government for evaluating and reimbursing university faculty (2000). The contributors argue that the new system places responsibility for compliance on the performer, not the checker. Thus, there is a shifting of responsibility that masks the underlying power dynamics: the indicator itself does the work of critique and the governed person seeks to conform to the terms of the government. Similar benefits devolve to treaty bodies that develop indicators: if the treaty body can persuade the country being governed to develop its own indicators, the committee can replace its practices of checking of country policies and actions with countries' self-checking (Rosga and Satterthwaite 2009). The turn to indicator creation marks a shift in the way the administration of human rights law takes place. Instead of pressuring countries to conform to human rights laws on the basis of ambiguous and contextualized accounts in country reports or case studies, reports in which each country is presented as shaped by its history, social structure, wealth, and political agendas, indicators provide comparable information in numerical terms. The burden of assessment rests on the indicator itself, with its agreed-upon standards and means of measurement. Although the experts developing one set of indicators for monitoring compliance with human rights conventions argued that the numbers were not to be used to rank or shame countries but only for assessing a country's progress over time, once an indicator has been created, such rankings are possible (Turku Report 2005, p. 7). The reliance on numbers, with their apparently simple and straightforward meanings, produces an unambiguous and easily replicated field for judgment. Compliance becomes far more open to inspection and assessment.

Moreover, responsibility for compliance shifts to the monitored organization, corporation, or country itself, which must not only seek to comply but monitor and report the success of its efforts. The enforcement body moves away from the role of an authority imposing criticisms to a body that registers performance in terms of already-established indicators. In other words, the process of assessing compliance shifts from the encounter between statements and rules in a quasi-judicial forum such as a treaty body hearing to the creation of the measure itself. Once the indicator has been established, compliance is simply a matter of recording performance according to the indicator. Treaty bodies are moving from asking countries to come up with their own

indicators toward a universal set of indicators for all countries which can be assessed impartially by the treaty body (Rosga and Satterthwaite 2009, p. 4). Corporations have clearly been active in defining the terms of the indicators by which their social responsibility will be judged.

In sum, the expansion of the use of indicators in global governance means that political struggles over what human rights or corporate social responsibility means and what constitutes compliance are submerged by technical questions of measurement, criteria, and data accessibility. Political debates about compliance shift to arguments about how to form an indicator, what should be measured, and what each measurement should represent. These debates typically rely on experts in the field of measurement and statistics, usually in consultation with experts in the substantive topic and in the national and international terrain. They build on previous research studies and knowledge generated by scholars. The outcomes appear as forms of knowledge, rather than as particular representations of a methodology and particular political decision about what to measure and what to call it. An indicator provides a transition from ambiguity to certainty, theory to fact, complex variation, and context to truthful, comparable numbers. In other words, the political process of judging and evaluating is transformed into a technical issue of measurement and counting by the diligent work of experts. Practices of measuring phenomena that are relatively easily counted, such as money or inventories of goods, are transplanted into domains far less amenable to quantification, such as frequency of torture or prevalence of ill-health. Technologies of knowledge developed in the economic domain move uneasily into these newer fields.

The creation of indicators reveals a slippage between the political and the technical. The slippage occurs in the way issues and problems are defined, in the identity and role of experts, in the relative power of the people engaged in producing and using indicators, and in the power and clout of the sponsoring organization. Through the apparatus of science and measurement, the indicator displaces judgment from governing bodies onto the indicator itself, which establishes standards for judgment. Nevertheless, indicators are inevitably political, rooted in particular conceptions of problems, and theories of responsibility. They represent the perspectives and frameworks of those who produce them as well as their political and financial power. What gets counted depends on which groups and organizations can afford to count. However, indicators differ significantly between those produced by a powerful organization, such as the World Bank, which scores and ranks countries, and more participatory processes such as the OHCHR human rights indicators in which the experts provide a framework, but the choice of indicators, methods, and data collection lies to a somewhat greater extent with the countries being measured.

The Genealogy of Indicators

Where did indicators come from? What is their genealogy? Since their creation in practices of financial management and governance, in Europe, perhaps four centuries ago, they have migrated across sectors and nations. The use of numerical information to understand the world reflects the creation of what Poovey (1998, p. xii) calls the "modern fact" as a form of knowledge. The modern fact is basic to the ways Westerners have come to know the world. It organizes most of the knowledge projects of the last four centuries (Poovey 1998, p. xiii). Numbers are the epitome of the modern fact because they seem to be simple descriptors of phenomena and to resist the biases of conjecture and theory, since they are subject to the invariable rules of mathematics. Numbers have become the bedrock of systematic knowledge because they seem free of interpretation, as neutral and descriptive. They are presented as objective, with an interpretive narrative attached to them by which they are given meaning. Numbers can be assigned to observed particulars in a way that makes them amenable to such manipulations and makes them amenable to a knowledge system that privileges quantity over quality and equivalence over difference (Poovey 1998, p. 4).

However, Poovey shows that numbers are not non-interpretive but embody theoretical assumptions about what should be counted, how to understand material reality, and how quantification contributes to systematic knowledge about the world (1998, p. xii). Establishing the understanding of numbers as an objective description of reality outside interpretation was a project of modernity. Although some see facts as interpreted, the idea that numbers guarantee value-free description is still pervasive (Poovey 1998, p. xxv). Poovey argues that the early nineteenth-century combination of numbers and analysis enabled professionals to develop systematic knowledge through non-interpretive descriptions. The nineteenth-century separation of numbers from interpretation made numbers different in kind from analytic accounts, locating them in a different stage in knowledge-producing projects. Since the numbers were different in kind from other knowledge, they could be developed by a special class of professionals who worked with them. Experts, professional knowledge-producers, took responsibility for managing this different kind of knowledge—knowledge that existed prior to policy and could be used in neutral ways to inform it (Poovey 1998, p. xv).

Statistics became increasingly important as a technology of governance in nineteenth-century Europe. As scholars of the intellectual history of statistics indicate, numbers as an instrument of knowledge production were developed first for business transactions, exemplified in particular by the invention of

double-entry bookkeeping, and subsequently as instruments of state governance (Poovey 1998). The use of numerical measures by states for administration and tax collection stretches back millennia, but it is only with the development of the modern state that statistics have been used to describe the characteristics of populations themselves. Quantification, with its aura of objectivity, became increasingly important to a variety of government and business functions in the nineteenth century, from developing cost-benefit measures for locating railroad lines to the need to measure life spans by life insurance companies in the mid-nineteenth century (Porter, 1995, pp. 106–121).

Contemporary global indicators inevitably rely on local data collection processes, although they may be created and managed at the international level. Local centers may understand the process differently, carry out the measurement tasks in different ways, or resist cooperating with national and international expectations. It is striking that all of the global governance indicator projects I have looked at are created in the global North, which sets the agenda, names the indicator, and assembles the criteria, while data collection typically takes place mostly in the global South. As the use of indicators enhances the exposure of nations to international scrutiny and potentially control, there may be forms of local resistance to the process.

Using Indicators for Governance

As tools of governance, indicators are commonly developed by powerful bodies seeking to manage and control populations or allocate resources. They may also be used to rank countries or organizations or to determine eligibility for a benefit. Indicators are not only directed at helping decision-makers decide where to build a railroad or in what country to invest, but also at promoting self-governance among the governed. By establishing standards according to which individuals, organizations, or nations should behave, indicators should inspire those who are measured to perform better and improve their ranking. Students in the USA are very familiar with the role that grades play in their educational lives. One of the reasons for creating indicators for treaty compliance is to promote nations taking steps to improve their performance according to the numerical standards of human rights treaties. Countries sometimes respond by emphasizing their status on indicators where they rank highly. For example, when Lithuania reported to the committee that monitors compliance with the Women's Convention, CEDAW on July 2, 2008, which I observed, the government representative, the Secretary of the Ministry of

Social Security and Labour, pointed out that according to the World Economic Forum's Report Global Gender Gap Index 2007, Lithuania was among the countries that made the most significant progress among the top 20 countries and now occupies 14th place. The minister also noted that Lithuania was in second place in the employment rate of women raising children below 12 years according to the EU Report on Gender Equality in 2008. Clearly, the minister was using these rankings to point out how well her country was succeeding in diminishing gender discrimination (CEDAW/C/LTU/Q/4).

The governed often shift their behavior in ways designed to improve their score, although they may do so in ways not desired by the producer of the indicator. As Rosga and Satterthwaite note, indicators have a relatively short life before those who are governed by them begin to change their behavior in order to enhance their score (2009). While this may be the desired outcome, it may also produce strategies to "game" the indicator. For example, some colleges graded down by the "US News and World Report" for low rates of alumni giving divide their alumnae gifts into three yearly payments. Although some highly ranked colleges have recently refused to participate at all, those ranked lower have relatively little power to challenge or change the system of ranking.

As indicators shift responsibility for governance from those in power to those who are governed, they may undermine autonomy, a sense of trust, and willingness to cooperate among certain kinds of populations. Strathern and her colleagues criticize the Research Assessment Exercise program of the British government which has introduced indicators of faculty productivity and activity as the basis for allocating revenues to academic departments (2000). As Strathern argues, this mechanism creates the standards to which universities then seek to govern themselves, but for professionals who work long hours with low pay under conditions of autonomy, this regime suggests a lack of trust and leads to alienation and resistance, producing exhaustion and withdrawal.

The turn to indicators is part of a new form of governance, one that engages the person in governing himself in terms of standards set by others. This new form of governance emphasizes "responsibilization", in which individuals are induced to take responsibility for their actions (O'Malley 1999). In some of the most successful examples, such as grades in school, the indicator comes to shape subjectivity, defining for the individual his or her degree of merit. These indicators promote self-management, what Rose and Miller call "government at a distance". He argues that new systems of governance have emerged in the post-war period that seek to control individual behavior through governance of the soul (Rose 1989, 1996, 1999).

In the liberal democracies of the post-war period, citizens are to regulate themselves, to become active participants in the process rather than objects of domination. Rose dates the formation of this self-managing system of governance to the 1950s but sees a major expansion during the era of neo-liberalism and the critique of the welfare state (Rose 1989, pp. 226–7). However, Kipnis (2008) criticizes Rose's emphasis on the connection of audit culture and neo-liberalism, since similar practices of monitoring occur in China under a very different political regime.

Indicator Governance and the Corporate Form

Indicators are a basic technology of corporate management and control, but as they move into the previously distinct domain of human rights and humanitarianism, the boundaries between business, the state, and what is commonly referred to as "civil society" blur. In practice, the corporation is increasingly intertwined with these other domains of society in discourse and in management strategy. The spread of its techniques of auditing and counting to the state and civil society is an instance of this seepage of the corporate form. Here, I will identify three forms of interchange.

The first is the donors' demand for performance evaluations of civil society organizations by foundations and governments. Social justice and humanitarian organizations face an increasingly onerous burden of quantifying their accomplishments, even when they are difficult to measure, and the data are expensive to produce, as discussed above. A further step in this direction is the US government's move to create indicator-based development funding. The Millennium Challenge Corporation (MCC), started in 2004, relies on competition among countries to allocate funding. Countries that perform better on the indicators established by the MCC are more likely to receive funding. This system replaces the earlier use of conditions that have to be met by countries receiving development aid. This approach emphasizes a country's responsibility for its governance and embodies the argument that effective government is fundamental to development.

The key concern of the MCC program is controlling corruption through promoting "good governance". Countries are measured by 17 indicators grouped into three broad categories: ruling justly, investing in people, and encouraging economic freedom. The indicators are all developed by other organizations. Five of the six governance indicators were developed by the World Bank while two are from Freedom House. Health and education indicators come from UNESCO and WHO, and economic freedom ones from

the World Bank and the Heritage Foundation's trade policy indicator. The MCC also uses the Corruption Perceptions Index of Transparency International and the US State Department Human Rights Report (2009). The process of selection involves four steps. The MCC Board identifies eligible countries from the low- and middle-income range, publishes the selection criteria, develops scorecards for each country, and, on the basis of these scorecards, selects some for assistance. Countries selected by the Board as eligible are invited to submit proposals for an MCC Compact. A few countries with a low score on one of the policy indicators are selected each year to participate in the Millennium Corporation Threshold Program to help raise their score and become eligible for a Millennium Challenge Grant. The Threshold Program is run by USAID (2009).

In a discussion of this program, in January 2008, at the American Enterprise Institute titled "Can Indicator-based Competition make Foreign Aid Work?", speakers emphasized that the turn to indicators is result of the emphasis on accountability. The overarching idea is to replace conditionalities with competition. Under this indicator approach, countries know what is expected of them and can compete for funds according to these standards. However, at this event, the representative from the UNDP said that he thought the mechanism was too complex, and conditions should be loosened (American Enterprise Institute [AEI] 2008). These examples suggest that work associated with the promotion of development, human rights, and good governance is increasingly being channeled by reliance on indicators.

The corporate form is also moving into domains of state and civil society governance with its engagement in processes of indicator development and data collection. Corporations are increasingly involved in the expensive and highly technical process of collecting and analyzing data and writing reports for NGOs, governments, and UN agencies. For example, a recent initiative of USAID East Africa and the USAID Inter-agency Gender Working group to develop a compendium of monitoring and evaluation indicators of violence against women and girls was developed by Monitoring and Evaluation to Assess and Use Results (MEASURE) Evaluation in collaboration with a technical advisory group of experts. The advisory group consisted of experts from United Nations Human Commissioner for Refugees (UNHCR), United States Agency for International Development (USAID), Center for Disease Control (CDC), United Nations Fund for Population Activities (UNFPA), World Health Organization (WHO), academics, independent consultants, and several people from MEASURE Evaluation, one of whom authored the report (Bloom 2008). MEASURE Evaluation describes itself as providing technical leadership

> through collaboration at local, national, and global levels to build the sustainable capacity of individuals and organizations to identify data needs, collect and analyze technically sound data, and use that data for health decision-making. We develop, implement and facilitate state of the art methods for and approaches to improving health information systems, monitoring and evaluation, and data use; and we collect, share, and disseminate information, knowledge, and best practices in order to increase the use of data and advance the field of health monitoring and evaluation in many countries. (MEASURE n.d.)

The organization is funded by USAID and works in partnership with the University of North Carolina, Tulane University, and ICF Macro, among others, revealing the collaboration of academic, government, and corporate actors (MEASURE n.d.).

ICF Macro is a large corporation that includes a program, MEASURE Demographic and Health Surveys (DHS), that since 1984 has provided technical assistance for 240 DHSs in 75 countries around the world. ICF Macro is based in the Washington, DC, area and maintains offices across the USA. It conducts projects for private and public sector clients in more than 125 countries. ICF Macro has annual revenues of approximately $150 million and more than 1100 employees, in 2009, joined ICF International (ICF 2017). Similarly, an Organisation for Economic Cooperation and Development (OECD) educational testing program, Programme for International Student Assessment (PISA), hired an international contractor, an Australian company, to work with each participating state to carry out the assessment. Student questionnaires and tests were developed by the international contractors, the PISA governing board, and functional expert groups (Bogdandy and Goldman 2009, p. 13). The development of data and analysis, and sometimes even the indicators themselves, is clearly a blend of public and private activity that brings together corporations, academics, NGOs, governments, and UN bodies as well as local, national, and international organizations. Data collection and analysis companies typically come from developed countries and often work in developing countries.

Not only are corporations increasingly involved in producing the data and measures that make up indicators used in the public domain but efforts to persuade corporations to be more socially responsible have also adopted this technology. As social movement activists, NGOs, the UN, and other NGOs seek to control the human rights, environmental, labor, and corruption practices of corporations, they have turned to the same strategies of governance that corporations exported to the social reformers. The emerging field of corporate social responsibility (CSR) relies on indicators of corporate

performance to assess companies (see Welker and Wood 2011). The UN's GC and the GRI are two of the most widely used global CSR systems—both rely on indicators to assess compliance with their general principles and both are voluntary. The GC website claims that it is the largest corporate citizenship initiative in the world. It says it launched the program in 2000 and as of May 2007 had more than 3000 companies from 100 countries, as well as over 700 civil society and international labor organizations, participating in the initiative. The GRI is an international network of business, civil society, labor, and professional institutions. This group has created a Reporting Framework through a consensus-seeking process. By 2006, more than 1000 organizations from nearly 60 countries had formally declared their use of the GRI Guidelines according to a GC report. GRI developed a set of detailed indicators which the GC adopted to implement its general principles.

The Global Compact Annual Review 2007 describes its monitoring process as a system of periodic reports by every signatory company every two years detailing its compliance with the GC's ten principles, articulated as indicators, plus its support for the MDGs. The ten principles cover human rights, labor, environmental issues, and corruption. The reports are called Communications on Progress (COPs). They should include statement of continued support for the GC by the Chief Executive Officer or other senior executive, description of practical actions of company's efforts to implement the GC principles and undertake partnership projects in support of broad UN goals, such as MDGs and measurements of expected outcomes, using indicators as much as possible or metrics such as the GRI guidelines. If a company fails to file a report within three years of signing or two years from its previous COP, it will be defined as inactive and dropped from the GC group (2007).

The 2008 Guidelines for COPs advocate presenting information about commitment, systems in place to insure compliance such as policies, programs, management systems, activities, and measures of outcomes. It recommends that reports should "Use performance indicators appropriate for your company's size, sector and unique operating environment, and also allow for benchmarking and comparability". In other words, companies are invited to develop their own metrics. "Companies should develop systems and evaluation programmes to assure that the information they are recording, collecting, analysing and disclosing is accurate and reliable. Importantly, this need not be a highly complex and expensive process, but could be as simple as a local Global Compact network peer review programme" (UN Global Compact 2008, p. 15). The reporting guidelines stress it is important to produce reliable

and specific measures in order to assess progress rather than to focus only on policies or activities. "Specific measurements that track actual performance are essential for ensuring continuous improvement" (UN Global Compact 2008, p. 17). Some of the internal benefits claimed for the process are discussion and awareness of these issues in the company while external benefits are enhancing the corporation's reputation (UN Global Compact 2008, p. 18). Thus, the CG represents another example of the mobilization of the argument that social responsibility is good for business, not just morality (Welker and Wood 2011).

The guidelines use over 30 indicators developed by GRI. Some focus on behavior, while others ask for numbers of training sessions or policies and management programs. The following list of illustrative indicators is characteristic of the GC approach of enumerating trainings or policies more than actual behavior.

HR 1 Percentage and total number of significant investment agreements that include human rights clauses or that underwent human rights screening.

HR 3 Total hours of employee training on policies and procedures concerning aspects of human rights that are relevant to operations, including the percentage of employees trained.

HR 4 Total number of incidents of discrimination and actions taken.

HR 5 Operations identified where the right to exercise freedom of association and collective bargaining may be at significant risk, and actions to support these rights.

HR 6 Operations identified as having significant risk for incidents of child labor and measures to contribute to eliminate child labor.

SO 5 Public policy positions and participation in public policy development and lobbying

SO 2 Percentage and total number of business units analyzed for risks related to corruption.

SO 3 Percentage of employees trained in organization's anti-corruption policies and procedures (UN Global Compact 2008, pp. 21, 33, 39).

The Guidelines suggest that companies check with their human resources, employee relations, supply management, legal, media and public relations, public affairs, or corporate relations offices for this information.

The GRI focuses on Sustainability Reporting Guidelines. In 2006, the organization published its third generation of guidelines, performance indicators, and indicator protocols, called GRI G3. The indicators developed for the

GRI can be used to address the ten principles of the GC. Although there are some differences, overall, the two voluntary reporting mechanisms cover roughly the same issues.

Thus, the monitoring system for GC and GRI is quite similar to that of UN treaty bodies, in which a governing organization confronts the dilemma of judging compliance based on information provided by the organization being judged. Like treaty body reports, the information requested focuses more on the existence of policies and training programs than on actual changes in behavior. Treaty bodies typically cope with this situation by politely asking for more information and focusing on information about laws and policies more than on data on performance. Nevertheless, treaty bodies constantly request more statistical data on outcomes and performance and are currently seeking to develop indicators for human rights. In both of these monitoring systems, indicators seem to offer a solution to the lack of independent information available to those who seek to govern.

Conclusions

In sum, indicators are a political technology that can be used for many different purposes, including advocacy, reform, control, and management. In some ways, indicators are like witchcraft. Witchcraft is the power to guide the flow of supernatural forces for good or harm. It is pervasive in societies that see supernatural forces as powerful actors in the world. Misfortunes and disease are the result of hostile supernatural forces, but healing and recovery from psychic and physical illness also rely on the mobilization of supernatural powers. Sometimes the same person is both a witch and a healer, since both depend on the ability to control these forces. Like witchcraft, indicators are a technology that exercises power, but in a variety of ways depending on who is using it for what purposes. And, like witchcraft, indicators presume a system of knowledge and a theory of how things happen that is hegemonic and rarely subjected to scrutiny, despite its critical role in the allocation of power.

As the world becomes ever more measured and tracked through indicators, it becomes increasingly important to sort out the technical and political dimensions of this new technology. Indicators produce readily understandable and convenient forms of knowledge about the world that shape the way policymakers as well as the general public understand the world. Those with long use have become naturalized as well as hegemonic, as in the case of grades for school performance. This is a form of knowledge production and governance that has expanded from its economic, corporate origins to a wide

array of uses in national and global governance. Indicators contribute to the calcification of categories, such as caste, race, or gender, which are subjected to categorical definition and measurement. The use of these statistical techniques, with their aura of certainty, is producing new knowledge of the social world and new opportunities for governance through self-governance. The expansion of indicator technology into new domains and spaces of governance is another way the corporate form is reshaping contemporary social life.

Notes

1. This document, reporting the discussion of an expert group meeting to develop an indicator for violence against women, convened by the United Nations Division for the Advancement of Women, the United Nations Economic Commission for Europe, and the United Nations Statistical Division, describes indicators as follows: "Indicators are part of the knowledge base needed to assist policy and decision-making. They help to raise awareness of an issue. Indicators, with their associated benchmarks, contribute to the monitoring of progress in achieving goals, and in policy evaluation. They enable an evidence-based comparison of trends over time, and within and between countries. Indicators on violence against women may also support the assessment of States' exercise of their due diligence obligation to prevent and address violence against women, and the effectiveness of related policies and other measures" (UN Expert Group Meeting Report, 8–10 October, p. 4).
2. Kaufman and Kraay emphasize the importance of sharing information on measurement error and the constituent elements of the indicator, but in their review of governance indicators they note that many indicators do not make this information available (2007).
3. The origins of the idea of the modern scientific fact and its representation by numbers, themselves subject to manipulation according to fixed rules, occurred along with the invention of double-entry bookkeeping as a mode of business management (Poovey 1998).

References

Alston, P. (2005). Richard Lillich Memorial Lecture: Promoting the Accountability of Members of the New UN Human Rights Council. *Journal of Transnational Law and Policy, 15*(1), 49–96.

American Enterprise Institute (AEI). (2008). *An Innovative Approach to Foreign Aid*. Retrieved from: https://www.aei.org/publication/an-innovative-approach-to-foreign-aid/

Bloom, S. S. (2008). *Violence Against Women and Girls: A Compendium of Monitoring and Evaluation Indicators*. Washington, DC: US AID.

Center for International Development, & Conflict Management (2017). *Research*. [Website] Retrieved from: https://cidcm.umd.edu/landing/Research

Cohn, B. S. (1996). *Colonialism and Its Forms of Knowledge: The British in India*. Princeton: Princeton University Press.

Davis, K. (2004). What Can the Rule of Law Variable Tell us About Rule of Law Reforms? *Michigan Journal of International Law, 26*, 141–161.

Davis, K. E., & Kruse, M. B. (2007). Taking the Measure of Law: The Case of the *Doing Business* Project. *Law and Social Inquiry, 32*(4), 1095–1120.

Dirks, N. B. (2001). *Castes of Mind: Colonialism and the Making of Modern India*. Princeton: Princeton University Press.

Doing Business. (2016). *Economy Rankings*. Washington, DC: Doing Business. Retrieved from: http://www.doingbusiness.org/EconomyRankings

Filmer-Wilson, E. (2005, March 31). *Summary Report of Material Collated Regarding Practical Guidance to Implementing Rights-Based Approaches, Human Rights Analyses for Poverty Reduction and Human Rights Benchmarks from Development Actors and Other Relevant Communities*. London: Department for International Development. Retrieved from: http://www.gsdrc.org/docs/open/con30.pdf

Goonesekere, S. (2004). Introduction: Indicators for Monitoring Implementation of CEDAW. In United Nations Development Fund for Women, South Asia Regional Office (Ed.), *CEDAW Indicators for South Asia: An Initiative*. Columbi: Centre for Women's Research (CENWOR).

Green, M. (2001). What We Talk About When We Talk About Indicators: Current Approaches to Human Rights Measurement. *Human Rights Quarterly, 23*, 1062–1097.

ICF. (2017). [Website]. Retrieved from: http://www.icf.com

Kaufmann, D., & Kraay, A. (2007). *Governance Indicators: Where Are We, Where Should We Be Going?* (Policy Research Working Paper, No. 4370). Washington, DC: The World Bank. Retrieved from: http://siteresources.worldbank.org/DEC/Resources/KraayKaufmannGovernanceIndicatorsSurveyNov12.pdf

Kipnis, A. B. (2008). Audit Cultures: Neoliberal Governmentality, Socialist Legacy, or Technologies of Governing? *American Ethnologist, 35*(2), 275–289.

Latour, B. (1987). *Science in Action: How to Follow Scientists and Engineers Through Society*. Cambridge, MA: Harvard University Press.

Malhotra, R. & Fasel, N. (2005). *Quantitative Human Rights Indicators: A Survey of Major Initiatives* (Background Paper for the UN Expert Meeting on Human Rights Indicators). Oslo: Norwegian Centre for Human Rights. Retrieved from: http://www.gaportal.org/sites/default/files/Quantitative%20Human%20Rights%20Indicators.pdf

Maurer, B. (2005). *Mutual Life, Limited: Islamic Banking, Alternative Currencies, Lateral Reason*. Princeton: Princeton University Press.

MEASURE Evaluation. (n.d.). *Mission*. [Website]. Retrieved from: https://www.measureevaluation.org/about/mission

Millennium Development Goals Indicators. (2017). [Website]. Retrieved from: https://mdgs.un.org/unsd/mdg/Default.aspx

O'Malley, P. (1999). Governmentality and the Risk Society. *Economy and Society, 28*, 138–148.

Poovey, M. (1998). *A History of the Modern Fact: Problems of Knowledge in the Sciences of Wealth and Society*. Chicago: University of Chicago Press.

Porter, T. M. (1995). *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*. Princeton: Princeton University Press.

Power, M. (1999). *The Audit Society: Rituals of Verification*. Oxford: Oxford University Press.

Randeria, S. (2006). Entangled Histories of Uneven Modernities: Civil Society, Caste Solidarities and Legal Pluralism in Post-Colonial India. In J. Keane (Ed.), *Civil Society – Berlin Perspectives* (pp. 213–242). Cambridge: Cambridge University Press.

Rose, N. (1989). *Governing the Soul: The Shaping of the Private Self*. London: Routledge.

Rose, N. (1996). The Death of the Social? Re-Figuring the Territory of Government. *Economy and Society, 25*, 327–356.

Rose, N. (1999). *Predicaments of Freedom*. Cambridge: Cambridge University Press.

Rosga, A. J., & Satterthwaite, M. L. (2009). The Trust in Indicators: Measuring Human Rights. *Berkley Journal of International Law (BJIL), 27*(2), 253–315.

Rottenburg, R. (2009). *Far-fetched Facts: A Parable of Development Aid*. Cambridge, MA: MIT Press.

Sen, A. K. (1999). *Development as Freedom*. New York: Anchor Books.

Shore, C., & Wright, S. (2000). Coercive Accountability: The Rise of Audit Culture in Higher Education. In M. Strathern (Ed.), *Audit Cultures: Anthropological Studies in Accountability, Ethics, and the Academy* (pp. 57–90). London/New York: Routledge.

Strathern, M. (Ed.). (2000). *Audit Cultures: Anthropological Studies in Accountability, Ethics, and the Academy*. London/New York: Routledge.

Turku Expert Meeting on Human Rights Indicators. (2005). [Website]. Retrieved from: www.abo.fi/instut/imr/indicators/index.html

UN Division for the Advancement of Women. (2007). *Report of the Expert Group Meeting: Indicators to Measure Violence Against Women*. Geneva: UN Division for the Advancement of Women, UN Economic Commission for Europe, UN Statistical Division.

UN Global Compact. (2007). *Annual Review*. New York: UN Global Compact. Retrieved from: https://www.unglobalcompact.org/docs/news_events/8.1/GCAnnualReview2007.pdf

UN Global Compact. (2008). *The Practical Guide to the United Nations Global Compact Communications on Progress (COPs). Revised Version of 2004 Report*. New York: United Nations Global Compact. Retrieved from: www.unglobalcompact.org

United Nations Development Programme (UNDP). (2000). Using Indicators for Human Rights in Accountability. In *Human Development Report 2000* (pp. 89–111). New York: UNDP.

US Agency of International Development. (2009). *Threshold Program*.

US State Department. (2009). *Human Rights Report 2009*. Washington, DC: US State Department. Retrieved from: https://www.state.gov/j/drl/rls/hrrpt/2009/

Von Bogdandy, A. & Goldman, M. (2009). *The Exercise of Public Authority Through National Policy Assessment* (Institute for International Law and Justice Working Paper 2009/2, Global Administrative Law Series). New York: Institute for International Law and Justice. Retrieved from: http://www.iilj.org/publications/the-exercise-of-international-public-authority-through-national-policy-assessment/

Welker, M., & Wood, D. (2011). Shareholder Activism and Alienation. *Current Anthropology, 52*(Suppl. 3), S57–S69.

22

Conclusions: Knowing and Governing

Debora Valentina Malito, Nehal Bhuta, and Gaby Umbach

This Handbook attempted to deliver a clearer understanding of the recent indicator culture (Merry 2016) by focusing on the production, consumption, and meta-consumption of metrics in the field of corruption, sustainability, and governance. The contributors to this Handbook did not all share the same assumptions on numbers and their narratives, and they hence did not come to the same conclusions. All chapters however provided a deeper understanding of the dialectical relationship between technology and social action, knowing and governing, and τέχνη (tèchne) and πολιτεία (politèia). The aim of this Handbook was hence to provide a wide theoretical and empirical state of the art on the use of indicators in global governance.

The Handbook's Part I collected interdisciplinary conceptual analyses, which frame the challenging puzzles of measuring and quantifying complex social, political, and economic phenomena. The various authors addressed theoretical and empirical aspects of indicators and their production and highlighted key conceptual, epistemological, and political problems that could be discerned. They addressed the ways in which indicators may be understood as exercises of power and governance at the global level and the potential limits of relying on them in the formulation of policy judgements and alternatives. As a result, this first part of the Handbook framed a set of critical questions

D.V. Malito (✉)
University of Cape Town, Cape Town, South Africa

N. Bhuta • G. Umbach
European University Institute, Florence, Italy

D.V. Malito et al. (eds.), *The Palgrave Handbook of Indicators in Global Governance*,
https://doi.org/10.1007/978-3-319-62707-6_22

about the reliability, utility, and malleability of indicators, which formed red threats throughout the rest of the Handbook.

Part II focused on the analytical questions related to the production of indicators, as well as several alternative measurement strategies. By collecting the experience of practitioners or scholars mostly associated with the production and use of the indicators analysed, this part provided institutional perspectives on the making of indicators. From the chapters in this second section, it is shown how conceptual and theoretical issues highlighted in Part I were resolved (or not resolved) in the real making of measures. The chapters generated insights into the ways in which internal institutional dynamics interact with scientific aspects to ultimately produce a certain kind of measure. They also explored how institutions use measures and to what extent "technical" aspects replace or do not replace "political" judgements within institutional decision-making.

Part III concentrated on analytical questions related to indicator use, as well as its implications and contestations. The third section followed some of these measures "into the field", bringing together multi-disciplinary case studies which examined how certain indicators have been used and how they have impacted on policy and politics in specific contexts. The individual case studies brought the insights and arguments of Parts I and II "down to the ground" by exploring how, in certain places at certain times, the use of indicators shaped public discussion, impacted on political judgements, and potentially altered political environments. The Handbook's third part hence provided a sense of the tangible power of indicators to shape governance at different scales—nationally, locally, and globally.

The introduction to this Handbook elaborated on the three topics that cut across its various chapters, that is, a growing post-metrological trend, the relevance of normativity, and the complex contextual power of indicators, and that will now be revisited in light of the insight generated by the contributions to this volume.

Post-metrological Realism

Indicators and indexes have increasingly become more sophisticated over the past decades. Many of the existing measures have tried to synthesise the multidimensionality of complex concepts such as governance or sustainability into single "best metrics". By doing so, they fell short of paying tribute to the concepts' inherent heterogeneity and the theoretical uncertainties embedded in them. Instead of exerting unifying or even convergence pressures, the

attempts at creating mono-dimensional representations of governance, corruption, or sustainability increased the number of measurement flaws and conceptual divides. Mirroring these difficulties, in this Handbook, scholars and practitioners criticised the conceptual challenges of "one-size-fits-all measurements" or of aggregate measures that are not entirely actionable and/or capable of providing "policy-relevant" information. Many contributions to this Handbook helped to track an important turn, that is characterised by the development of disaggregate, micro-level, local experience-based measures that may not be comparable across countries but that are better suited to guide reforms, monitoring, and decision-making "on the ground". Based on a reflexive criticism of the existing cross-country, aggregate approaches, many providers have therefore elaborated different pathways.

Many authors also elaborated on the blind spots of the current state of the art. Alonso and Garcimartin (Chap. 4) conceptualised good governance as the result of the quality of institutions (capacity, predictability, adaptability, and credibility). They focused on disaggregate information "by fields and function of the state" as the fundamental building blocks to enable collective action. However, most measures especially of governance still employed inputs, as well the "capacity" or "quality" of government. Adding to this analysis, Rotberg (Chap. 2) pointed out that governance should hence be defined in terms of performance across five bundles of political goods (security, rule of law, participation, economic sustainability, and human development).

While Raghunandan (Chap. 17) explored the virtues of investigative and auditing approaches as opposed to quantification in measuring corruption in India, Ivanyna and Shah (Chap. 18) proposed a citizen-centric approach based on public opinion surveys focusing on governance. For Greenwood (Chap. 14), the future of sustainable development measures required stronger attention to all capital stocks (namely public infrastructure, natural resource capital, human capital, and social capital), which contributed to the production of well-being. Evaluating the oil price-based fiscal rule introduced in Nigeria, Obinyeluaku (Chap. 13) illustrated why developing fiscal indicators for African countries heavily dependent on volatile revenues is considered a pressing policy problem.

Many practitioners that contributed to this Handbook emphasised the importance of mapping the procedural, micro-level, disaggregate dimension of governance. Lafortune et al. (Chap. 9) explained the virtues of "dashboard approaches", such as the one employed by the OECD. This methodology included "narrowly defined and often detailed indicators on government inputs, activities, outputs and outcomes" (p. 209). It provided disaggregate measures of governance and information on a variety of conditions (political,

legal/administrative, cultural, and socio-economic), "without creating any super-aggregation" or "super-composites" aimed at "summarising the performance of activities (e.g. digital government performance, regulatory governance) or government as a whole into a single figure" (p. 209).

Recanatini (Chap. 8) elaborated on the concept of within- and cross-country heterogeneity to highlight the importance of micro-level data for thorough measurement. Central to the World Bank's anti-corruption strategy was, hence, the incorporation of citizen participation into the design of strategies and reforms. Through country-specific surveys of households, business communities, and public officials, this new approach advocated for a participatory monitoring process in data analysis and reform strategies. By focusing on the usually underestimated process of implementation, the World Bank aimed to address a perceived gap between de jure and de facto systems of governance. Instead of providing general guidelines for measurement, the Bank possibly bridged the gap between controllers and controlled, that is, between those that are subjects and those that are objects of measurement and reforms.

Berti (Chap. 10) revealed how a more detailed analysis of country-specific factors has been essential for the European Commission to formulate policy recommendations on fiscal policy. Classical indicators (such the debt to GDP or GNI ratio) that exclusively centred on the solvency criterion in the past raised many criticisms. Practitioners had recently embraced different perspectives, and the European Commission's approach to measuring fiscal sustainability had evolved over the last decades: from an original focus on long-term tax gap indicators, the Commission adopted a multi-dimensional approach, redefining existing indicators and creating new ones. As a result, the sustainability of public finances was now defined as the solvency of the public sector in the context of short-, medium- and long-term challenges.

But not only "experienced" providers, such as the OECD, the World Bank, or the European Union, have undergone processes of reflexive critique. Relatively "new producers", such as Global Integrity and the World Justice Project (WJP), have internalised criticisms of their first generation of measures and redesigned the conceptual frameworks and methodological platforms on which their own measurement tools were based. Dougherty et al. (Chap. 11) presented an interesting participatory exercise by the WJP related to their rule of law index. The index combined expert opinion with citizens' perceptions and opinion polls. This participatory approach allowed the WJP to define its epistemic framework as "broadly accepted" (p. 269), even though they admit that "a combination of sources, instruments and methods are

essential in attempting to capture a broad view of the rule of law in a nation" (p. 269).

Feigenblatt and Tonn (Chap. 12) elaborated on the constant development process of the Global Integrity report: country coverage, design of indicators, and data gathering have been modified in the recent past to create more plural measures. "In construing the indicators, Global Integrity makes an effort to work with commonly accepted views, best practice and international benchmarks, but understands that in many cases there will still be alternative views" (p. 285). To allow space for more reflexive critique, Global Integrity transferred the final judgement on the objectivity of their measures directly to the users, improving the transparency and public availability of their methodological choices, "so that users can ultimately determine for themselves whether the data prove useful for their measurement objectives" (p. 286).

Normativity

As claimed by Alonso and Garcimartin (Chap. 4), scholars do not generally agree whether governance should be separated from (Fukuyama 2013) or bound together (Rothstein 2011) with normative dimensions. Measuring governance hence reflects such dualities, and contributors to this Handbook agreed that normativity matters. Yet, by questioning whether and how normativity should, or should not, inform the process of measuring governance, they provided different answers.

For Alonso and Garcimartin, existing measures fail to deliver on their diagnostic function because of a normative deficit: reflections on the normative frameworks and institutions that condition development and governance have been limited. "[B]efore developing new indicators we should define more carefully the theoretical basis used to build them and to select variables" (p. 90). In contrast to this, Rotberg (Chap. 2) claimed that existing measures were flawed exactly because they were more normative than descriptive. He proposed a conceptualisation of governance in terms of performance instead. "Citizens everywhere—in every culture, in every political jurisdiction" (p. 35) share the same expectations from the government. Trying "hard not to be prescriptive or normative" (p. 41), this approach normalises a particular perception of governance that ignores the respective historical experience or normative frame that influenced its formation.

For many authors in the Handbook, indicators however hold an intrinsic normative quality, often framed in terms of quasi-neutral and natural change. Beschel (Chap. 7) claimed that a strong normative dimension is either implicit

or explicit in using indicators "to create impetus for positive change …The goal is not just to accurately reflect reality, but to create incentives" (p. 165). This idea of "positive changes" and "incentives" was a common theme among practitioners to justify their policy-oriented commitment to measurement. Elaborating on the World Bank's governance and anti-corruption reforms, Recanatini (Chap. 8) pointed out that "the system of incentives" regulating the relationships between citizens and the government is central to any practical perspective of change (p. 186). For Feigenblatt and Tonn (Chap. 12), Global Integrity's new measurement strategy was expressively built on a theory of change anchored in the role of adaptive learning for fostering open governance. What is more, although Global Integrity had traditionally avoided "rigid prescriptive approaches", the authors emphasised the need for more consistent, actionable indicators and claimed that "indicators that fail to explicitly unpack and define certain concepts become too fuzzy for consistent coding—and therefore diminish the value of a final product that proposes to offer users practical value" (p. 285). For Dougherty et al. (Chap. 11), methodological procedures like data collection and weighting depended on the purposes, users, and consumers of modern indicators. There was indeed "a tension among the needs of various users of information. While all of them seek timely and accurate information, they each have different goals and emphases" (p. 263). Accountability, actionability, adaptability, comparability, flexibility, and simplicity are some of the qualities that a wide range of users (from the business and academic community to development practitioners and institutions) were looking for.

The contributions to this Handbook also tended to connect the normativity of indicators to broader transformations that altered the nature of authority and economic production in the contemporary world order. The measurement space "is certainly not a space of pure objectivity", claimed Morse (p. 2), "[d]espite the quantitative nature" of sustainability indicators and "the maths used to create them". Yet, this normative terrain was not a homogeneous or unified one. Sustainable development was on its own "a multiverse with many dimensions" (p. 432); indicators were "human constructs to help meet the perceived (by humans) needs of humans". Different institutions and practitioners provided different standards for developing sustainable indicators, each of which might have particular interests to promote, but they did not necessarily overlap or correspond to each other. The result was a hence "complex and diverse 'ecosystem' [of sustainability indicators], some of which overlap in terms of their focus and some share the same data; some evolve over time and some have even been discontinued" (p. 432). For Malito (Chap. 5), indicators of governance and stateness embedded and

reiterated specific, neo-liberal policy prescriptions. Yet, a dialectical coexistence of interests and prescriptions existed among institutional providers that informed the production stage. The "creative disorder" of measuring governance reflected the hybridity and pervasiveness of the neo-liberal paradigm of governance; as a result, development prescriptions often even coexisted with neo-trusteeship logics. For Urueña (Chap. 16), a measure of corruption such the Corruption Perception Index (CPI) set the tone of "the epistemic ground on which the normative and policy debates about corruption take place" (p. 384). A factual narrative about the existence of corruption was hence necessary to justify the involvement of a number of global actors and mechanisms in the anti-corruption normative regime. Such evidence was not built upon hard data, such as criminal prosecutions, but rather on perceptions that corruption was an obstacle to a certain policy-oriented idea of development. "This is the empty space that the CPI fills", concluded Urueña, creating an image of corruption that is visible only through "the prism" created by Transparency International (p. 384).

In her analysis of indicators in the field of human rights compliance, Merry (Chap. 21) focused on the "slippage between the political and the technical", where corporate management penetrated and shaped the domains of state and civil society governance as well social life. The production and use of indicators not only contributed to amplify the Global North-South divide. It also contributed to the "calcification of categories, such as caste, race or gender, which are subjected to categorical definition and measurements" (p. 488).

The Contextual Power of Indicators

As became evident throughout the chapters of this Handbook, measuring is an inherent political process, and indicators are instruments of governance in itself. To some extent, they substitute governance processes by influencing decision-making through forms of soft power. In some cases, they directly reconfigure political relations, creating priorities and influencing judgements. Indicators hence play an important role in defining multiple relationships of power and dominance. They are mutant creatures, used in different contexts with a variety of purposes. Implications for power relations are numerous, and the power of indicators therefore remains highly contextual and dependent on how they are used and enrolled in particular circumstances. By focusing on some of the most debated measuring fields, such as corruption and sustainability, authors in this Handbook provide an interesting overview of the multiple directions of the "power by indicators".

Although the conceptual and symbolic role of indicators in knowledge production is widely acknowledged, many analysts claim that corruption indicators so far have had very limited power to influence national policy dimensions or to be incorporated into specific policy proposals. Indicators rather contributed to the generation of new ontologies of the real. Urueña (Chap. 16) interpreted Transparency's International "activism through numbers" as focused on mastering the reality indispensable to enhance an anti-corruption normative regime as a neutralised field, one that only accidentally "coincides" with neo-liberal development policy. For Cooley, the new ontology generated by international rankings did therefore not necessarily enhance our understanding of global governance challenges. In the case of corruption, international rankings reiterated a distorted image of corruption as a phenomenon bounded within "descrete... national units" (p. 51). To borrow from Urueña, what was globally visible as corruption was only visible from the national "prism", created by measures such as CPI. How "transnational forces facilitate the relevant transactions in a globalised world" (p. 62) thus remained largely obscured.

The impact or influence of indicators on the ground hence remains highly contextualised because indicators are also prone to a variety of political and instrumental uses. Indicators are malleable creatures, and they can also justify policy orientations that have not been originally enshrined in them. Musaraj (Chap. 15) illustrated how a corruption perception survey such as the one of USAID/IDRA (United States Agency for International Development/Institute for Development Research and Alternatives) played an important role in the post-socialist transformations of a country. The survey was extensively used by local actors as well as US representatives to influence national policy-making and justice reforms in Albania. Raghunandan (Chap. 17) illustrated how indicators also prompted other externalities: civil society initiatives to measure corruption in India enabled not only the collection of feedback on the quality and adequacy of public services but also the diffusion of auditing approaches.

The contributions to this Handbook furthermore helped to understand how measures of fiscal sustainability have been used to reduce and remove complexity (Bhuta et al. 2014). In her analysis of fiscal sustainability metrics, Greenwood (Chap. 14) highlighted as to what extent most metrics oversimplify complex issues related to national income and economic capacity. The most important implication was that incomplete measures provided wrong conceptual frameworks for policy-making. Also, Blot (Chap. 20) claimed that the capacity of the most commonly used indicators to formulate fiscal policy recommendations was limited by partial equilibrium assumptions that do not consider other macro-economic conditions. The optimal level for public debts established by the EU (the 60% debt ratio) represented more of an institutional value than of a theoretically well-grounded one. Also analysing the use

of fiscal sustainability indicators by global newspapers, Morse (Chap. 19) revealed that indicators are malleable and dynamic, often used to reduce complexity: indicators were often "presented as *de facto* 'black boxes' and, understandably, their consumers treat them in that way and take it on trust that they 'work'" (p. 446). Interestingly, concluded Morse, this was not really "a debate about indicators *per se*" but rather about "the competing visions of what can be regarded desirable, of which they are but expressions; at the end of the day it is the ideas that matter and not the indicators" (p. 446).

Conclusions

Measuring is a way of doing politics. Scholars and practitioners, who contributed to this Handbook, agreed that measuring mattered because of its instrumentality in governing. They however came to contrasting conclusions about the way forward and the volume therefore reflected this variety of discordant interpretations.

Some contributions contested the effectiveness of the existing indicator culture, and addressed methodological solutions inspired by standards of scientific objectivity, like a focus on performance indicators (Chap. 2) or institutional quality dimensions (Chap. 4), improved techniques, and internal validity and reliability (Chaps. 13, 14, 17, and 18). Other were critical about the contemporary indicator culture because of normative premises, as well as intended and unintended consequences, ranging from the creation of distorted ontologies of the real (Chaps. 3 and 16), and the instrumental, conceptual hybridity in serving the pervasiveness of the neo-liberal paradigm of governance (Chap. 5) and the simplification of social complexity (Chaps. 6, 13 and 20), over the external interference in the legitimacy of domestic policy decision-making (Chap. 15), to the dissemination of a "corporate form of thinking and governance into broader social spheres" (Chap. 21, p. 478). Many contributors proposed solutions that in their view better suited the decentralisation of governance. Their chapters did not demand pure mechanical objectivity but rather a better transformation of politics into metrics through locally embedded (Chap. 14), disaggregated (Chaps. 9 and 10), micro-level (Chap. 8), country-specific (Chap. 17) data and systems of knowledge creation.

This Handbook tried to create some kind of order in the avalanche of indicators surrounding us. It did so by tracing some developments, trends, and trajectories about the diagnostic, normative, and power dimensions of indicators. Many aspects remain in need of further research, for which this volume presented an empirically informed analytical foundation.

References

Bhuta, N., Malito, D. V., & Umbach, G. (2014). *Representing, Reducing or Removing Complexity: Indicators of Sustainability and Fiscal Sustainability* (Robert Schuman Centre for Advanced Studies Working Paper, 2014/78) Florence: European University Institute. Retrieved from: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2462023

Fukuyama, F. (2013). What is Governance? *Governance, 26*(3), 347–368.

Merry, S. E. (2016). *The Seductions of Quantification: Measuring Human Rights, Gender Violence, and Sex Trafficking*. Chicago: University of Chicago Press.

Rothstein, B. (2011). *The Quality of Government: Corruption, Social Trust, and Inequality in International Perspective*. Chicago: University of Chicago Press.

Index[1]

[1] Note: Page numbers followed by 'n' refers to note.

D.V. Malito et al. (eds.), *The Palgrave Handbook of Indicators in Global Governance*,
https://doi.org/10.1007/978-3-319-62707-6