76778

Governance Indicators: Where Are We, Where Should We Be Going?

Daniel Kaufmann • Aart Kraay

Progress in measuring governance is assessed using a simple framework that distinguishes between indicators that measure formal rules and indicators that measure the practical application or outcomes of these rules. The analysis calls attention to the strengths and weaknesses of both types of indicators as well as the complementarities between them. It distinguishes between the views of experts and the results of surveys and assesses the merits of aggregate as opposed to individual governance indicators. Some simple principles are identified to guide the use and refinement of existing governance indicators and the development of future indicators. These include transparently disclosing and accounting for the margins of error in all indicators, drawing from a diversity of indicators and exploiting complementarities among them, submitting all indicators to rigorous public and academic scrutiny, and being realistic in expectations of future indicators. JEL codes: H1, O17

> Not everything that can be counted counts, and not everything that counts can be counted. —Albert Einstein

Most scholars, policymakers, aid donors, and aid recipients recognize that good governance is a fundamental ingredient of sustained economic development. This growing understanding, initially informed by a very limited set of empirical measures of governance, has spurred intense interest in developing more refined, nuanced, and policy-relevant indicators of governance. This article reviews progress in measuring governance, emphasizing empirical measures explicitly designed to be comparable across countries and in most cases over time. The goal is to provide a structure for thinking about the strengths and weaknesses of different types of governance indicators that can inform both the use of existing indicators and ongoing efforts to improve them and develop new ones.[1]

The first section of this article reviews definitions of governance. Although there are many broad definitions of governance, the degree of definitional disagreement

can easily be overstated. Most definitions appropriately emphasize the importance of a capable state that is accountable to citizens and operating under the rule of law. Broad principles of governance along these lines are naturally not amenable to direct observation and thus to direct measurement. As Albert Einstein noted, "Not everything that counts can be counted." Many different types of data provide information on the extent to which these principles of governance are observed across countries. An important corollary is that any particular indicator of governance can usefully be interpreted as an imperfect proxy for some unobserved broad dimension of governance. This interpretation emphasizes throughout this review a recurrent theme that there is measurement error in all governance indicators, which should be explicitly considered when using these kinds of data to draw conclusions about cross-country differences or trends in governance over time.

The second section addresses what is measured. The discussion highlights the distinction between indicators that measure specific rules "on the books" and indicators that measure particular governance outcomes "on the ground." Rules on the books codify details of the constitutional, legal, or regulatory environment; the existence or absence of specific agencies, such as anticorruption commissions or independent auditors; and so forth—components intended to provide the key *de jure* foundations of governance. On-the-ground measures assess *de facto* governance outcomes that result from the application of these rules (Do firms find the regulatory environment cumbersome? Do households believe the police are corrupt?). An important message in this section concerns the shared limitations of indicators of both rules and outcomes: Outcome-based indicators of governance can be difficult to link back to specific policy interventions, and the links from easy-to-measure *de jure* indicators of rules to governance outcomes of interest are not yet well understood and in some cases appear tenuous at best. They remind us of the need to respect Einstein's dictum that "not everything that can be counted counts."

The third section examines whose views should be relied on. Indicators based on the views of various types of experts are distinguished from survey-based indicators that capture the views of large samples of firms and individuals. A category of aggregate indicators that combine, organize, provide structure, and summarize information from these different types of respondents is examined. The fourth section examines the rationale for such aggregate indicators, and their strengths and weaknesses.

The set of indicators discussed in this survey is intended to provide leading examples of major governance indicators rather than an exhaustive stocktaking of existing indicators in this taxonomy.[2] A feature of efforts to measure governance is the preponderance of indicators focused on measuring *de facto* governance outcomes and the paucity of measures of *de jure* rules. Almost by necessity, *de jure* rules-based indicators of governance reflect the views or judgments of experts. In

contrast, the much larger body of *de facto* indicators captures the views of both experts and survey respondents.

The article concludes with a discussion of the way forward in measuring governance in a manner that can be useful to policymakers. The emphasis is on the importance of consumers and producers of governance indicators clearly recognizing and disclosing the pervasive measurement error in any type of governance indicators. This section also notes the importance of moving away from oft-heard false dichotomies, such as "subjective" or "objective" indicators or aggregate or disaggregated ones. For good reason, virtually all measures of governance involve a degree of subjective judgment, and different levels of aggregation are appropriate for different types of analysis. In any case, the choice is not either one or the other, as most aggregate indicators can readily be unbundled into their constituent components.

What Does *Governance* Mean?

The concept of governance is not a new one. Early discussions go back to at least 400 BCE, to the *Arthashastra*, a treatise on governance attributed to Kautilya, thought to be the chief minister to the king of India. Kautilya presents key pillars of the "art of governance," emphasizing justice, ethics, and anti-autocratic tendencies. He identifies the duty of the king to protect the wealth of the state and its subjects and to enhance, maintain, and safeguard this wealth as well as the interests of the kingdom's subjects.

Despite the long provenance of the concept, no strong consensus has formed around a single definition of governance or institutional quality. For this reason, throughout this article the terms *governance, institutions*, and *institutional quality* are used interchangeably, if somewhat imprecisely. Researchers and organizations have produced a wide array of definitions. Some definitions are so broad that they cover almost anything (such as the definition "rules, enforcement mechanisms, and organizations" offered in the World Bank's *World Development Report 2002: Building Institutions for Markets*). Others, like the definition suggested by North (2000), are not only broad but risk making the links from good governance to development almost tautological: "How do we account for poverty in the midst of plenty? . . . We must create incentives for people to invest in more efficient technology, increase their skills, and organize efficient markets . . . . Such incentives are embodied in institutions."

Some of the governance indicators surveyed capture a wide range of development outcomes. While it is difficult to draw a line between governance and the ultimate development outcomes of interest, it is useful at both the definitional and measurement stages to emphasize concepts of governance that are at least somewhat removed from development outcomes themselves. An early and narrower

definition of public sector governance proposed by the World Bank is that "governance is the manner in which power is exercised in the management of a country's economic and social resources for development" (World Bank 1992, p. 1). This definition remains almost unchanged in the Bank's 2007 governance and anticorruption strategy, with governance defined as "the manner in which public officials and institutions acquire and exercise the authority to shape public policy and provide public goods and services" (World Bank 2007, p. 1).

Kaufmann, Kraay, and Zoido-Lobatón (1999a, p. 1) define governance as "the traditions and institutions by which authority in a country is exercised. This includes the process by which governments are selected, monitored and replaced; the capacity of the government to effectively formulate and implement sound policies; and the respect of citizens and the state for the institutions that govern economic and social interactions among them."

Although the number of definitions of governance is large, there is some consensus. Most definitions agree on the importance of a capable state operating under the rule of law. Interestingly, comparing the last three definitions cited above, the one substantive difference has to do with the explicit degree of emphasis on the role of democratic accountability of governments to their citizens. Even these narrower definitions remain sufficiently broad that there is scope for a wide diversity of empirical measures of various dimensions of good governance.

The gravity of the issues dealt with in these definitions of governance suggests that measurement is important. In recent years there has been debate over whether such broad notions of governance can be usefully measured. Many indicators can shed light on various dimensions of governance. However, given the breadth of the concepts, and in many cases their inherent unobservability, no single indicator or combination of indicators can provide a completely reliable measure of any of these dimensions of governance. Rather, it is useful to think of the various specific indicators discussed below as all providing imperfect signals of fundamentally unobservable concepts of governance. This interpretation emphasizes the importance of taking into account as explicitly as possible the inevitable resulting measurement error in all indicators of governance when analyzing and interpreting any such measure. As shown below, however, the fact that such margins of error are finite and still allow for meaningful country comparisons across space and time suggests that measuring governance is both feasible and informative.

Governance Rules or Governance Outcomes?

This section examines both the rules-based and outcome-based indicators of governance. A rules-based indicator of corruption might measure whether countries have legislation prohibiting corruption or have an anticorruption agency.

An outcome-based measure could assess whether the laws are enforced or the anticorruption agency is undermined by political interference. The views of firms, individuals, nongovernmental organizations (NGOs), or commercial risk-rating agencies could also be solicited regarding the prevalence of corruption in the public sector. To measure public sector accountability, one could observe the rules regarding the presence of formal elections, financial disclosure requirements for public servants, and the like. One could also assess the extent to which these rules operate in practice by surveying respondents regarding the functioning of the institutions of democratic accountability.

Because a clear line does not always distinguish the two types of indicators, it is more useful to think of ordering different indicators along a continuum, with one end corresponding to rules and the other to ultimate governance outcomes of interest. Because both types of indicators have their own strengths and weaknesses, all indicators should be thought of as imperfect, but complementary proxies for the aspects of governance they purport to measure.

Rules-Based Indicators of Governance

Several rules-based indicators are used to assess governance (tables 1 and 2). They include the Doing Business project of the World Bank, which reports detailed information on the legal and regulatory environment in a large set of countries; the Database of Political Institutions, constructed by World Bank researchers, and the POLITY-IV database of the University of Maryland, both of which report detailed factual information on features of countries' political systems; and the Global Integrity Index (GII), which provides detailed information on the legal framework governing public sector accountability and transparency in a sample of 41 countries, most of them developing economies.

At first glance, one of the main virtues of indicators of rules is their clarity. It is straightforward to ascertain whether a country has a presidential or a parliamentary system of government or whether a country has a legally independent anticorruption commission. In principle, it is also straightforward to document details of the legal and regulatory environment, such as how many legal steps are required to register a business or fire a worker. This clarity also implies that it is straightforward to measure progress on such indicators. Has an anticorruption commission been established? Have business entry regulations been streamlined? Has a legal requirement for disclosure of budget documents been passed? This clarity has made such indicators very appealing to aid donors interested in linking aid with performance indicators and in monitoring progress on such indicators. Set against these advantages are three main drawbacks.

They are less "objective" than they appear. It is easy to overstate the clarity and objectivity of rules-based measures of governance. In practice, a good deal of

Table 1. Sources and Types of Information Used in Governance Indicators

| | Type of indicator | | | |
| | Rules-based | | Outcomes-based | |
Source of information	Broad	Specific	Broad	Specific
Experts				
Lawyers		DB		
Commercial risk-rating agencies			DRI, EIU, PRS	
Nongovernmental organizations		GII	HER, RSF, CIR, FRH	GII, OBI
Governments and multilaterals			CPIA	PEFA
Academics	DPI, PIV		DPI, PIV	
Survey respondents				
Firms				ICA, GCS, WCY
Individuals			AFR, LBO, GWP	
Aggregate indicators combining experts and survey respondents			TI, WGI, MOI	

Note: AFR is Afrobarometer, CIR is Cingranelli-Richards Human Rights Dataset, CPIA is Country Policy and Institutional Assessment, DB is Doing Business, DPI is Database of Political Institutions, DRI is Global Insight DRI, EIU is Economist Intelligence Unit, FRH is Freedom House, GCS is Global Competitiveness Survey, GII is Global Integrity Index, GWP is Gallup World Poll, HER is Heritage Foundation, ICA is Investment Climate Assessment, LBO is Latinobarómetro, MOI is Ibrahim Index of African Governance, OBI is Open Budget Index, PEFA is Public Expenditure and Financial Accountability, PIV is Polity IV, PRS is Political Risk Services, RSF is Reporters Without Borders, TI is Transparency International, WCY is World Competitiveness Yearbook, and WGI is Worldwide Governance Indicators.

Source: Authors' compilation based on data from sources listed in table 2.

subjective judgment is involved in codifying all but the most basic and obvious features of a country's constitutional, legal, and regulatory environments. (It is no accident that the views of lawyers, on which many of these indicators are based, are commonly referred to as *opinions*.) In Kenya in 2007, for example, a constitutional right to access to information faced being undermined or offset entirely by an official secrecy act and by pending approval and implementation of the Freedom of Information Act. In this case, codifying even the legal right to access to information requires careful judgment as to the net effect of potentially conflicting laws. Of course, this drawback of ambiguity is not unique to rules-based measures of governance: interpreting outcome-based indicators of governance can also involve ambiguity, as discussed below. There has been less recognition, however, of the extent to which rules-based indicators also reflect subjective judgment.

The links between indicators and outcomes are complex, possibly subject to long lags, and often not well understood. These problems complicate the interpretation of rules-based indicators. In the case of rules-based measures, some of the most basic features of countries' constitutional arrangements have little normative content on their own; such indicators are for the most part descriptive. It makes

Table 2. Country Coverage and Frequency of Governance Surveys

Name	Number of countries covered	Frequency of survey	Web site
Afrobarometer	18	Triennial	www.afrobarometer.org
Cingranelli-Richards Human Rights Dataset	192	Annual	www.humanrightsdata.com
Country Policy and Institutional Assessment	136	Annual	www.worldbank.org
Doing Business	175	Annual	www.doingbusiness.org
Database of Political Institutions	178	Annual	http://econ.worldbank.org
Global Insight DRI	117	Quarterly	www.globalinsight.com
Economist Intelligence Unit	120	Quarterly	www.eiu.com
Freedom House	192	Annual	www.freedomhouse.org
Global Competitiveness Survey	117	Annual	www.weforum.org
Global Integrity Index	41	Triennial	www.globalintegrity.org
Gallup World Poll	131	Annual	www.gallupworldpoll.com
Heritage Foundation	161	Annual	www.heritage.org
Investment Climate Assessment	94	Irregular	www.investmentclimate.org
Latinobarómetro	17	Annual	www.latinobarometro.org
Ibrahim Index of African Governance	48	Triennial	www.moibrahimfoundation.org
Open Budget Index	59	Annual	www.openbudgetindex.org
Polity IV	161	Annual	www.cidcm.umd.edu/polity/
Political Risk Services	140	Monthly	www.prsgroup.com
Public Expenditure and Financial Accountability	42	Irregular	www.pefa.org
Reporters without Borders	165	Annual	www.rsf.org
World Competitiveness Yearbook	47	Annual	www.imd.ch

Source: Authors' compilation.

little sense, for example, to presuppose that presidential (as opposed to parliamentary) systems or majoritarian (as opposed to proportional) representation in voting arrangements are intrinsically good or bad. Interest in such variables as indicators of governance rests on the case that they may matter for outcomes, often in complex ways. In their influential book, Persson, Torsten, and Tabellini (2005) document how these features of constitutional rules influence the political process and ultimately outcomes such as the level, composition, and cyclicality of public spending (Acemoglu 2006) challenges the robustness of these findings). In such cases, the usefulness of rules-based indicators as measures of governance depends crucially on how strong the empirical links are between such rules and the ultimate outcomes of interest.

Perhaps the more common is the less extreme case in which rules-based indicators of governance have normative content on their own, but the relative

importance of different rules for outcomes of interest is unclear. The GII, for example, provides information on the existence of dozens of rules, ranging from the legal right to freedom of speech to the existence of an independent ombudsman to the presence of legislation prohibiting the offering or acceptance of bribes. The Open Budget Index (OBI) provides detailed information on the budget processes, including the types of information provided in budget documents, public access to budget documents, and the interaction between executive and legislative branches in the budget process. Many of these indicators arguably have normative value on their own: having public access to budget documents is desirable and having streamlined business registration procedures is better than not having them.

This profusion of detail in rules-based indicators leads to two related difficulties in using them to design and monitor governance reforms. The first is that as a result of absence of good information on the links between changes in specific rules or procedures and outcomes of interest, it is difficult to know which rules should be reformed and in what order. Will establishing an anticorruption commission or passing legislation outlawing bribery have any impact on reducing corruption? If so, which is more important? Should, instead, more efforts be put into ensuring that existing laws and regulations are implemented or that there is greater transparency, access to information, or media freedom? How soon should one expect to see the impacts of these interventions? Given that governments typically operate with limited political capital to implement reforms, these trade-offs and lags are important.

The second difficulty in designing or monitoring reforms arises when aid donors or governments set performance indicators for governance reforms. Performance indicators based on changing specific rules, such as the passage of a particular piece of legislation or a reform of a specific budget procedure, can be very attractive because of their clarity: it is straightforward to verify whether the specified policy action has been taken.[3] Yet "actionable" indicators are not necessarily also "action worthy," in the sense of having a significant impact on the outcomes of interest. Moreover, excessive emphasis on registering improvements on rules-based indicators of governance leads to risks of "teaching to the test" or, worse "reform illusion," in which specific rules or procedures are changed in isolation with the sole purpose of showing progress on the specific indicators used by aid donors.

Major gaps exist between statutory rules on the books and their implementation on the ground. To take an extreme example, in all 41 countries covered by the 2006 GII, accepting a bribe is codified as illegal, and all but three countries (Brazil, Lebanon, and Liberia) have anticorruption commissions or similar agencies. Yet there is enormous variation in perceptions-based measures of corruption across these countries. The 41 countries covered by the GII include the Democratic

Republic of Congo, which ranks 200 out of 207 countries on the 2006 Worldwide Governance Indicators (WGI) control of corruption indicator, and the United States, which ranks 23.

Another example of the gap between rules and implementation (documented in more detail in Kaufmann, Kraay, and Mastruzzi 2005) compares the statutory ease of establishing a business with a survey-based measure of firms' perceptions of the ease of starting a business across a large sample of countries. In industrial countries, where *de jure* rules are often implemented as intended, the two measures correspond quite closely. In contrast, in developing economies, where there are often gaps between *de jure* rules and their *de facto* implementation, the correlation between the two is very weak; the *de jure* codification of the rules and regulations required to start a business is not a good predictor of the actual constraints reported by firms. Unsurprisingly, much of the difference between the *de jure* and *de facto* measures could be statistically explained by *de facto* measures of corruption, which subverts the fair application of rules on the books.

The three drawbacks—the inevitable role of judgment even in "objective" indicators, the complexity and lack of knowledge regarding the links from rules to outcomes of interest, and the gap between rules on the books and their implementation on the ground—suggest that although rules-based governance indicators provide valuable information, they are insufficient on their own for measuring governance. Rules-based measures need to be complemented by and used in conjunction with outcome-based indicators of governance.

Outcome-Based Governance Indicators

Most indicators of governance are outcome based, and several rules-based indicators of governance also provide complementary outcome-based measures. The GII, for example, pairs indicators of the existence of various rules and procedures with indicators of their effectiveness in practice. The Database of Political Institutions measures not only such constitutional rules as the presence of a parliamentary system, but also outcomes of the electoral process, such as the extent to which one party controls different branches of government and the fraction of votes received by the president. The Polity-IV database records a number of outcomes, including the effective constraints on the power of the executive.

The remaining outcome-based indicators range from the highly specific to the quite general. The OBI reports data on more than 100 indicators of the budget process, ranging from whether budget documentation contains details of assumptions underlying macroeconomic forecasts to documentation of budget outcomes relative to budget plans. Other less specific sources include the Public Expenditure and Financial Accountability indicators, constructed by aid donors with inputs of recipient countries, and several large cross-country surveys of firms—including

the Investment Climate Assessments of the World Bank, the Executive Opinion Survey of the World Economic Forum, and the World Competitiveness Yearbook of the Institute for Management Development—that ask firms detailed questions about their interactions with the state.

Examples of more general assessments of broad areas of governance include ratings provided by several commercial sources, including Political Risk Services, the Economist Intelligence Unit, and Global Insight–DRI. Political Risk Services rates 10 areas that can be identified with governance, such as "democratic accountability," "government stability," "law and order," and "corruption." Large cross-country surveys of individuals such as the Afrobarometer and Latinobarómetro surveys and the Gallup World Poll ask general questions, such as "Is corruption widespread throughout the government in this country?"

The main advantage of outcome-based indicators is that they capture the views of relevant stakeholders, who take actions based on these views. Governments, analysts, researchers, and decisionmakers should, and often do, care about public views on the prevalence of corruption, the fairness of elections, the quality of service delivery, and many other governance outcomes. Outcome-based governance indicators provide direct information on the *de facto* outcome of how *de jure* rules are implemented.

Outcome-based measures also have some significant limitations. Such measures, particularly where they are general, can be difficult to link back to specific policy interventions that might influence governance outcomes. This is the mirror image of the problem discussed above: Rules-based indicators of governance can also be difficult to relate to outcomes of interest. A related difficulty is that outcome-based governance indicators may be too close to ultimate development outcomes of interest. To take an extreme example, the Ibrahim Index of African Governance includes a number of ultimate development outcomes, such as per capita GDP (gross domestic product), growth of GDP, inflation, infant mortality, and inequality. While such development outcomes are surely worth monitoring, including them in an index of governance risks making the links from governance to development tautological.

Another difficulty has to do with interpreting the units in which outcomes are measured. Rules-based indicators have the virtue of clarity: either a particular rule exists or it does not. Outcome-based indicators by contrast are often measured on somewhat arbitrary scales. For example, a survey question might ask respondents to rate the quality of public services on a five-point scale, with the distinction between different scores left unclear and up to the respondent.[4] In contrast, the usefulness of outcome-based indicators is greatly enhanced by the extent to which the criteria for differing scores are clearly documented. The World Bank's Country Performance and Institutional Assessment (CPIA) and the Freedom House indicators are good examples of outcome-based indicators based

on expert assessments that provide documentation of the criteria used to assign specific scores on the indicators they compile. In the case of surveys, questions can be designed to ensure that responses are easier to interpret: rather than asking respondents whether they think "corruption is widespread," respondents can be asked whether they have been solicited for a bribe in the past month.

An example illustrates some of the main advantages and disadvantages of the two types of measures. Figure 1 compares alternative indicators of democratic accountability, a key dimension of governance. The horizontal axis measures a very broad outcome-based indicator, taken from the 2005 Voice of the People survey, a large cross-country household survey (www.voice-of-the-people.net). It asks households to indicate whether they think elections in their country are free and fair. The vertical axis reports two indicators of the quality of electoral institutions, taken from Global Integrity. The points labeled "*de jure*" are based on a factual assessment of the existence of a number of specific institutions related to

Figure 1. *De facto* and *de jure* Indicators of Elections
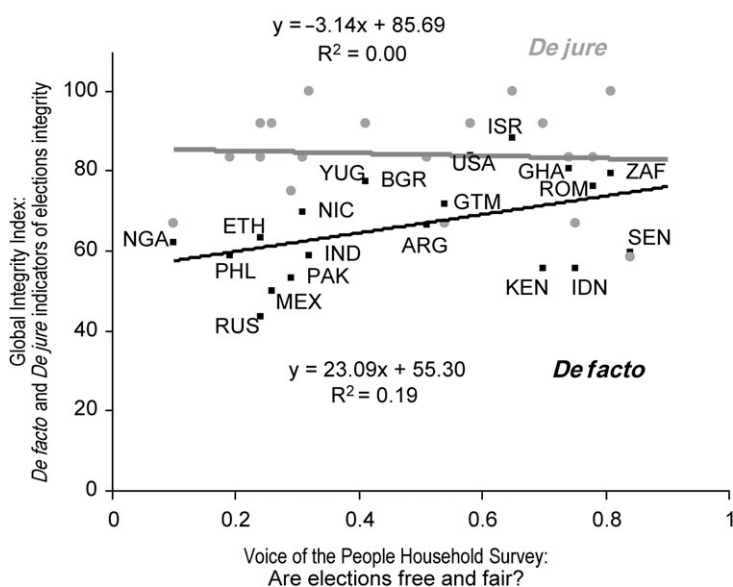
Note: ARG is Argentina, ARM is Armenia, AZE is Azerbaijan, BEN is Benin, BRA is Brazil, BGR is Bulgaria, ZAR is Democratic Republic of Congo, EGY is Egypt, ETH is Ethiopia, GEO is Georgia, GHA is Ghana, GTM is Guatemala, IND is India, IDN is Indonesia, ISR is Israel, KEN is Kenya, KGZ is Kyrgyz Republic, LBN is Lebanon, LBR is Liberia, MEX is Mexico, MNP is Montenegro, MOZ is Mozambique, NPL is Nepal, NIC is Nicaragua, NGA is Nigeria, PAK is Pakistan, PHL is Philippines, ROM is Romania, RUS is Russia, SEN is Senegal, YUG is Serbia, SLE is Sierra Leone, ZAF is South Africa, SDN is Sudan, TJK is Tajikistan, TZA is Tanzania, UGA is Uganda, USA is United States, VNM is Vietnam, YEM is Yemen, and ZWE is Zimbabwe.

Source: Authors' analysis based on data described in the text.

elections, such as the existence of a legal right to universal suffrage and the existence of an election monitoring agency. The points labeled "*de facto*" capture the assessment of Global Integrity's experts as to the effectiveness of these institutions.[5]

Several messages emerge from this figure. First, in some cases rules-based measures of governance show remarkably little variation across countries, with all countries receiving scores close to 100, indicating perfect scores on the *de jure* basis of this important aspect of governance. As of 2005, for example, every country surveyed by Global Integrity promised the legal right to vote, and a statutorily independent election-monitoring agency existed in all but three countries (Lebanon, Montenegro, and Mozambique). Second, the links between a specific objective indicator of rules and the broad outcome of interest (citizens' satisfaction with elections) is at best very weak, with a correlation between the two measures that is in fact slightly negative. Third, outcome-based indicators explicitly focusing on the *de facto* implementation of rules can be useful. A noteworthy feature of Global Integrity is its pairing of indicators of specific rules with assessments of their functioning in practice. The correlation of the *de facto* measure with the broad outcome measure of interest taken from the Voice of the People survey is much stronger (0.46) than the correlation with the *de jure* measure. The correlation is far from perfect, however, indicating the importance of relying on a variety of indicators when assessing governance in a country.

Whose Views Should We Rely On?

A variety of governance assessments are produced by experts on behalf of commercial risk-rating agencies and NGOs. The GII and the OBI, for example, rely on locally recruited experts in each country to complete their detailed questionnaires about governance, subject to peer review. Commercial organizations such as the Economist Intelligence Unit rely on a network of local correspondents in a large set of countries to provide information underlying the ratings they produce. Other advocacy organizations, such as Amnesty International, Freedom House, and Reporters without Borders, also rely on networks of respondents for the information underlying their assessments.

Governments and multilateral organizations are also major producers of expert assessments. Some of the most notable include the Country Policy and Institutional Assessments, produced by the World Bank, the African Development Bank, and the Asian Development Bank. Each of these assessments is based on the responses of each institution's country economists to a detailed questionnaire, responses that are then reviewed for consistency and comparability across countries. The Public

Expenditure and Financial Accountability indicators mentioned earlier are also based on experts' views.

Several large cross-country surveys of firms and individuals contain questions on governance. These include the Investment Climate Assessment and the Business Environment and Enterprise Performance Surveys conducted by the World Bank; the Executive Opinion Survey of the World Economic Forum; the World Competitiveness Yearbook; Voice of the People; and the Gallup World Poll.

Expert Assessments

Expert assessments have several major advantages, which account for their preponderance among various types of governance indicators. One is cost: it is much less expensive to ask a selection of country economists at the World Bank to provide responses to a questionnaire on governance as part of the CPIA process than to carry out representative surveys of firms or households in a hundred or more countries. The second advantage is that expert assessments can more readily be tailored for cross-country comparability: Many of the organizations listed in table 2 have elaborate benchmarking systems to ensure that scores are comparable across countries. Finally, for certain aspects of governance experts are the natural respondents for the type of information being sought. (Consider, for example, the OBI's detailed questionnaire on national budget processes, the particulars of which are not the sort of common knowledge that survey data can easily collect.)

Expert assessments nevertheless have several important limitations. A basic one is that, like survey respondents, different experts may have different views about similar aspects of governance. While this is not surprising, it suggests that users of governance indicators should be cautious about relying too heavily on any one set of expert assessments. These differences are evident in comparing the CPIA ratings of the World Bank and the African Development Bank, which in recent years harmonized their procedures for constructing CPIA ratings. An identical questionnaire covering 16 dimensions of policy and institutional performance is completed by two very similar sets of expert respondents—country economists with in-depth experience working on behalf of these two organizations in the countries they are assessing. Despite the homogeneity of the respondents and the very similar rating criteria, there are nontrivial differences between the two organizations' assessments on the 16 components of the CPIA (table 3). For example, the 0.67 correlation between the two assessments on the question on transparency, accountability, and corruption in the public sector is far from perfect, suggesting that it is prudent to base assessments of governance for policy purposes on the views of a variety of expert assessments.[6]

Table 3. Correlation Among Alternative Indicators of Corruption

| | Expert assessments | | | | Surveys | |
Indicator	World Bank CPIA	African Development Bank CPIA	Global Integrity	World Markets Online	World Economic Forum Executive Opinion Survey	Gallup World Poll
World Bank CPIA	1.00	0.67	0.30	0.56	0.25	0.13
African Development Bank CPIA		1.00	0.49	0.51	0.45	0.24
Global Integrity			1.00	0.34	0.29	0.11
World Markets Online				1.00	0.88	0.59
World Economic Forum Executive Opinion Survey					1.00	0.70
Gallup World Poll						1.00

Source: Authors' analysis based on data described in the text.

The second criticism that the country ratings assigned by different groups of experts are too highly correlated is just the opposite. Suppose that one set of experts comes up with an assessment of governance for a set of countries based on its own independent research and the second set of experts simply reproduces the assessments of the first. In this case, the high correlation of two expert assessments cannot be interpreted as evidence of their accuracy. Rather, it would reflect the fact that the two sources make correlated errors in measuring governance.[7]

Nevertheless, even if the errors made by two data sources are highly, but not perfectly correlated, there will be benefits to relying on both data sources. The important empirical question is whether this hypothetical correlation of errors across sources is large or not. Empirically identifying correlations in errors across sources is difficult. Simply observing whether the assessments provided in the two data sources are highly correlated is not enough, as the high correlation can reflect the fact that both sources are either measuring governance accurately or making correlated measurement errors.

To make progress, one needs to make identifying assumptions. Kaufmann, Kraay, and Mastruzzi (2006) detail two sets of assumptions that allow potential sources of correlation in the errors to be disentangled. One is that surveys of firms or individuals are less likely to make errors that are correlated with other data sources than, for example, assessments by commercial risk-rating agencies. If this is the case, however, one would expect that the assessments of commercial risk-rating agencies would be very highly correlated with one another, but less so with surveys. This turns out not to be the case. The average correlation of the five

major commercial risk-rating agencies for corruption in 2002–05 was 0.80. The correlation of each of these assessments with a large cross-country survey of firms was slightly higher (0.81), in contrast with what one would expect if the rating agencies had correlated errors. Conducting this exercise across all six aggregate governance indicators reveals at most modest evidence of error correlation. While this is unlikely to be the final word on this important question, it is a useful step forward to propose and implement tests of error correlation based on explicit identifying assumptions.

The third criticism is that expert assessments are subject to various biases. Some researchers claim that many of these sources are biased toward the views of the business community, which may have very different views of what constitutes good governance than do other types of respondents. In short, goes the critique, businesspeople like low taxes and less regulation, while the public good demands reasonable taxation and appropriate regulation. This critique does not seem particularly compelling. If it were true, the responses of commercial risk-rating agencies, which serve mostly business clients, or the views of firms themselves to questions about governance, should not be highly correlated with ratings provided by respondents who are more likely to sympathize with the common good, such as individuals, NGOs, or public sector organizations. Yet, in most cases, these correlations are strong (Kaufmann, Kraay, and Mastruzzi 2007b). Cross-country surveys of firms and of individuals, such as the World Economic Forum's Executive Opinion Survey and the Gallup World Poll, yield similar corruption rankings, with the two surveys correlated at 0.7 (table 3).

Another potential source of bias in expert assessments, particularly those produced by NGOs, is that they are colored by the ideological orientation of the ratings organization. Kaufmann, Kraay, and Mastruzzi (2004) find that the assessments of think tanks and firm surveys are not systematically correlated with the political orientation of a country's government, casting doubt on this possible source of bias. A potentially greater problem of bias is at the country respondent level. For example, the views of pro-government and antigovernment "experts" might be very different, affecting both levels and trends over time. This risk is perhaps greatest for the sources that rely on local experts, such as the GII. This risk is also much more difficult to test for systematically, as the biases may affect individual country scores without introducing systematic biases into the source as a whole. Nevertheless, careful comparisons of many different data sources can often turn up anomalies in a single source that require more careful scrutiny.

Surveys of Firms and Individuals

Governance indicators derived from surveys of firms and individuals have the fundamental advantage that they elicit the views of the ultimate beneficiaries of good

governance, citizens and firms in a country. The views of these stakeholders matter because they are likely to act on those views. If firms or individuals believe that the courts and the police are corrupt, they are unlikely to try to use their services (Hellman and Kaufmann 2004). Individuals are less likely to vote or to hold their elected leaders accountable if they think that elections are not free and fair.

Another advantage of governance indicators based on surveys of domestic firms and individuals is greater domestic political credibility. Governments often dismiss external expert assessments of governance as uninformed pontification by outsiders. It is much harder for them to dismiss the views of their own citizens or of firms operating in their country. Survey-based data on governance can therefore be particularly useful in galvanizing the politics of governance reforms. The experience of many countries implementing their own in-depth Governance and Anti-Corruption diagnostics (assisted by the World Bank Institute and other agencies and implemented with institutions in the requesting country), based on in-country surveys of enterprises, users of services, and public officials, supports this point: the views expressed by thousands of domestic stakeholders provide powerful input for action to reformist policymakers and civil society groups.

Set against these important advantages of surveys are a number of disadvantages. First, there is the usual array of potential problems with any type of survey data, ranging from issues of sampling design to issues of nonresponse bias. Expert assessments, which are based on the views of a very small number of respondents, are less likely to be representative of the population of firms or households.[8] While these generic issues are important for all surveys, the focus here is on difficulties specific to measuring governance using survey data.

Some survey questions on governance can be vague and open to interpretation. An interesting example comes from the innovative recent work by Razafindrakoto and Roubaud (2006). They use specially designed surveys in eight African countries to contrast corruption perceptions based on household surveys with those based on expert assessments. The unique feature of this exercise is that the experts were asked to predict the country-level average responses from the household survey. Experts' ratings were essentially uncorrelated with the household survey responses. The authors conclude that the household surveys capture the "objective reality" of petty corruption and that the experts are just plain wrong.

Their interpretation that there is measurement error only in the expert assessment and not in the household survey is contestable. Households were asked whether they had been "victims of corruption." There are a variety of reasons why households might falsely think they were victimized by corruption. For example, a patient waiting in line to see a state-provided doctor might think (incorrectly) that people at the head of the line had bribed someone to get there. Conversely, households might well have paid a bribe, received the associated benefit, and found themselves quite satisfied and not at all "victimized" by the transaction. A more modest

interpretation of their finding is that there likely is measurement error in both the household survey and the matching expert assessments. Moreover, in many other cases, expert assessments and household survey responses are strongly correlated across much larger samples of countries.

Well-designed survey questions on corruption have become increasingly specific. For example, questions in the Executive Opinion Survey of the World Economic Forum in some years have asked firms to specifically report the fraction of contract value solicited in bribes on public procurement contracts. Greater attention is also being paid to techniques that enable respondents to report more truthfully to sensitive questions. For example, questions about corruption put to firms are often prefaced by "in your experience, do firms like your own typically pay bribes for....?" Innovative techniques such as randomized response methods are used to protect the confidentiality of individual responses by allowing respondents to "camouflage" their response to sensitive questions by generating some of their responses at random based on the outcome of a coin toss, although these methods have not yet been widely used in large cross-country surveys.[9] A related concern has to do with surveys in authoritarian countries, where respondents might legitimately be fearful of responding truthfully to any question that might be interpreted as critical of the government.

Another potential difficulty in cross-country surveys is cultural bias. It is often argued that because respondents in different countries may have different norms regarding what does or does not constitute corruption, their responses are not comparable across countries. Presumably, however, these cultural biases should not be present in cross-country expert assessments that are deliberately designed to be comparable across countries. Moreover, in many cases it turns out that surveys and expert assessments tend to produce very similar cross-country rankings. Kaufmann, Kraay, and Mastruzzi (2006) document strong correlations between expert assessments and the World Economic Forum's Executive Opinion Survey for six different dimensions of governance. A glance at table 3 provides similar examples: the cross-country correlation between the corruption assessments of World Markets Online, a commercial rating agency, and the Executive Opinion Survey is 0.88. While culture undoubtedly matters in interpreting survey responses across countries, the problem does not appear to be a first-order difficulty.

In short, each type of data has its own strengths and weaknesses. As neither type of respondent is clearly superior for all purposes, it important to rely on a diversity of data sources.

Should Aggregate or Individual Indicators Be Used?

Does it make sense to combine individual indicators of governance into aggregate or composite indicators by combining information from multiple sources? Table 1

includes three aggregate indicators, the WGI, the Corruption Perceptions Index (CPI) of Transparency International, and the very recently released Ibrahim Index of African Governance.

The WGI consist of six aggregate indicators of governance covering more than 200 countries and combining cross-country data on governance provided by 30 organizations. The CPI measures only corruption, using a smaller set of data drawn from nine organizations. The WGI control of corruption indicator uses these nine data sources, as well as 13 others not used in the CPI. The Ibrahim Index is an extremely broad collection of a variety of types of governance indicators and several very broad development outcomes, including per capita income, growth, inequality, and poverty. This makes the Ibrahim Index by far the broadest indicator surveyed here, but it also makes it difficult to think of it as a pure governance indicator, because it contains many development outcomes as well. However, three of the five components of the Ibrahim Index—based primarily on subjective governance measures, such as those used by Transparency International and the WGI—correspond more closely to established notions of governance.

Ubiquitous Measurement Error

All governance indicators have limitations, which make them imperfect proxies for the concepts they are intended to measure. The presence of measurement error in all governance indicators that this implies is central to the rationale for constructing aggregate indicators. It is useful to distinguish between two broad types of measurement error that affect all types of governance indicators.

Any specific governance indicator will have measurement error relative to the concept it seeks to measure, because of intrinsic measurement challenges. A survey question about corruption, for example, will have the usual sampling error associated with it. Efforts to objectively document the specifics of the institutional environment or regulatory regime face challenges in coming up with a factually accurate description of the relevant laws and regulations in each setting. Measures of the composition and volatility of public spending, for example, which are sometimes interpreted as indicators of undesirable policy instability, are subject to all of the usual difficulties in measuring public spending consistently across countries and over time. Finally, different groups of experts may come up with different assessments of the same phenomenon in a particular country. These divergences of opinion can also be interpreted as measurement error.

To the extent that one is interested in broad concepts of governance, any specific indicator is almost by definition an imperfect measure of the broader concepts to which it pertains, no matter how accurate or reliable it is. A specific assessment of corruption in public procurement would not be fully informative

about overall corruption in the public sphere even if it were fully accurate about this specific type of corruption. Information about the statutory requirements for business entry regulation need not reflect the actual practice of how these requirements are implemented on the ground, and they are not informative about regulatory burdens in other areas. Information about freedom of the press is only one of many factors contributing to the accountability of governments to their citizens. Notwithstanding some clear advantages that specificity of an indicator may have for some purposes, one should be careful not to interpret them as sufficient statistics for broader notions of governance.

How important is this measurement error? Unfortunately, the vast majority of governance indicators do not explicitly acknowledge the extent of measurement error. One of the few exceptions is the WGI, discussed below. Fortunately, some simple calculations can shed light on the likely magnitude of measurement error in individual governance indicators as well. The key to doing so is to identify pairs of indicators that measure similar concepts, up to an unavoidable measurement error component. A useful way to interpret the imperfect correlation between the World Bank's CPIA and African Development Bank's CPIA regarding transparency and corruption is to note that both are measuring the same concept of transparency and corruption but with a degree of measurement error. Intuitively, the less measurement error there is in these two sources, the more correlated they should be. Thus one can interpret the correlation between them as saying something about the degree of measurement error present.

More formally, think of the observed scores from two organizations, y_1 and y_2, as a combination of a signal of unobserved governance, g, and source-specific noise, ε_1 and ε_2 (that is, $y_1 = g + \varepsilon_1$ and $y_2 = g + \varepsilon_2$). Assume that the variance of measurement error in the assessments of the two organizations is the same, and without loss of generality, that the variance of governance is one.[10] Some simple arithmetic reveals that the standard deviation of measurement error is $SD(\varepsilon) = \sqrt{(1 - \rho)/\rho}$, where ρ is the correlation between the two expert assessments.[11] For several pairs of indicators discussed in the article, this standard deviation of error ranges from 0.70 to 1.53 (table 4). The standard errors associated with standard deviation of error indicators such as the WGI are much smaller, reflecting the benefits of aggregation in reducing noise in the individual indicators. The standard error for the WGI estimate of control of corruption for a typical country in 2006 is just 0.17, or less than a quarter of the standard error of the most precise pair of individual indicators in this example.

To appreciate the magnitude of this measurement error, it is useful to go one step further and calculate the width of a 90 percent confidence interval for governance based on any one of these individual indicators and on the additional assumption that governance and the error term are jointly normally distributed. The width of the confidence interval is $2x1.64xSD(g|y) = 3.28x\sqrt{1 - p}$.

Table 4. Measurement Error in Individual Governance Indicators

Measure	Correlation	Standard deviation of error	Width of confidence interval
Transparency, accountability, and corruption	0.67	0.70	1.88
World Bank CPIA-16			
African Development Bank CPIA-16			
Business entry regulation	0.48	1.04	2.37
World Economic Forum Executive Opinion Survey			
Doing Business			
Elections	0.30	1.53	2.74
Global Corruption Barometer Survey			
Global Integrity Elections Index			

Source: Authors' analysis based on data described in the text.

Since the assumptions imply that 95 percent of countries would have governance levels between −2 and 2, these figures imply that a 90 percent confidence interval for governance for any individual country would span one-half to two-thirds of the entire most likely range of the governance indicator.

Why Aggregate Indicators?

All indicators of governance include measurement error. Aggregate indicators of governance can be a useful way of combining, organizing, and summarizing information from different sources, thereby reducing the influence of measurement error in any individual indicator. Aggregation also allows for the construction of explicit margins of error for both the aggregate indicator and its component individual indicators.

The WGI illustrate how these margins of error can be calculated (box 1). The statistical methodology underpinning the WGI (the unobserved-components model) explicitly assumes that the true level of governance is unobservable and that the observed empirical indicators of governance provide imperfect signals of the fundamentally unobservable concept of governance. This formalizes the notion that all available indicators are imperfect proxies for governance. The estimates of governance that come out of this model are simply the conditional expectation of governance in each country, conditioning on the observed data for each country. Moreover, the unobserved-components model allows one to summarize uncertainty about these estimates for each country with the standard deviation of unobserved governance, conditional on the observed data. These standard deviations can be used to construct confidence intervals for governance estimates, often referred to informally as margins of error. Intuitively, the larger the number of data sources available for a given country, the smaller these

margins of error should be. The variance of the error term can be estimated in each individual underlying governance indicator using this methodology, following a calculation that generalizes the simple one discussed above.

Box 1. The Worldwide Governance Indicators: Critiques and Responses

---

The Worldwide Governance Indicators (WGI) are among the most widely used cross-country governance indicators (see Kaufmann, Kraay, and Mastruzzi 2007a for a description). They report on six dimensions of governance for more than 200 countries for 1996–2006. The indicators are based on hundreds of underlying individual indicators drawn from 30 organizations, based on responses from tens of thousands of citizens, enterprise managers, and experts.

As one of the most prominent and widely used collections of cross-country governance indicators, the WGI have naturally generated criticism. Most of these criticisms appear largely invalid (Kaufmann, Kraay, and Mastruzzi 2007b).

Lack of Comparability over Time and across Countries
Several critics have raised concerns that the WGI are not comparable over time and across countries, that the indicators use units that set the global average of governance to be identical in all periods, that comparisons of pairs of countries or single countries over time are based on different sets of underlying data sources, and that there are substantial margins of error in the aggregate WGI.

These criticisms appear unjustified, for several reasons. First, there is no clear evidence of a trend in one direction or another in global averages of governance in any of the underlying individual data sources (the overall evidence pointing to general stagnation). The choice of a constant global average is therefore no more than an innocuous choice of units. Second, changes in the set of underlying data sources on average contributes only minimally to changes over time in countries' scores on the aggregate WGI; the majority of cross-country comparisons using the aggregate WGI are based on a substantial number of common data sources. Third, the presence of explicit margins of error in the WGI is an important advantage, serving as a useful antidote to superficial comparisons of country ranks or country performance over time that are often made with other governance indicators. A substantial fraction of cross-country and over-time comparisons using the WGI result in statistically significant differences, suggesting that the WGI are informative.

Biases in Expert Assessments
Several critics have alleged various biases in the data sources underlying the WGI, including an excessive emphasis on business-friendly regulation on the part of some data providers; ideological biases, such as a bias against left-wing governments, on the part of some data providers; and "halo effects," whereby countries with good economic performance receive better-than-warranted governance scores. Convincing empirical evidence in support of such biases has not been provided. Empirical work by Kaufmann, Kraay, and Mastruzzi (discussed in the main text) suggests that these biases are quantitatively unimportant.

Correlated Perception Errors
Several critics have suggested that expert assessments make similar errors when assessing the same country, leading to correlations in the perception errors across expert assessments. While this is plausible, there is little convincing empirical evidence; work by Kaufmann, Kraay, and Mastruzzi (discussed below) suggests that these biases are quantitatively unimportant.

Box 1. Continued

A related concern is that correlated perception errors will lead to the overweighting of such sources in the aggregate WGI, which weights individual data sources by estimates of their precision, which in turn are based on the observed intercorrelation among sources. Given the at best modest evidence of correlated perceptions errors, this is unlikely to be quantitatively important. The WGI country rankings are highly robust to alternative weighting schemes (Kaufmann, Kraay, and Mastruzzi 2006).

Definitional Issues

Some critics have taken issue with the definitions of governance and thus the assignment of individual governance indicators to the six aggregate WGI. As there is no consensus on the definition of governance, there cannot be any right or wrong definitions or corresponding measures of governance. That said, most reasonable definitions of governance cover similar broad areas, and aggregate indicators capturing these broad areas are likely to be similar. Moreover, as virtually all of the individual indicators underlying the WGI are publicly available on the WGI Web site, researchers can easily construct alternative indicators corresponding to their preferred notions of governance.

Reliance on "Subjective" Data

Various critics have argued that the perceptions-based data on which the WGI are based do no more than reflect vague and generic perceptions rather than specific objective realities and that "specific, objective, and actionable" measures of governance are needed to make progress in governance reforms. Virtually all governance indicators necessarily involve some element of subjectivity. Perceptions-based data are extremely valuable, because they capture the views of relevant stakeholders who act on these views. Moreover, the links from specific changes to policy rules are very difficult to link to changes in outcomes of interest, making it is difficult to identify indicators that are "action worthy" as opposed to merely "actionable."

From the standpoint of users the margins of error associated with estimates of governance are nontrivial. For many pairs of countries with similar scores on the 2006 WGI control of corruption indicator, the confidence intervals overlap, indicating that the small differences between them are unlikely to be statistically, or practically, significant (figure 2). However, possible pair-wise comparisons between countries do result in significant differences. Roughly two-thirds of the possible pair-wise comparisons of corruption across countries result in differences that are significant at the 90 percent confidence level, and nearly three-quarters of comparisons are significant at the 75 percent confidence level. Clearly, far fewer pair-wise comparisons would be significant if they were based on any single individual indicator, whose margins of error had not been reduced by averaging across alternative data sources. For the WGI control of corruption indicator, for example, only 16 percent of cross-country comparisons based on a typical individual data source, such as Global Insight-DRI, would be significant at the 90 percent confidence level.

The WGI are unusual among governance indicators in their transparent recognition of such margins of error. The vast majority of investment climate and governance indicators simply report country scores or ranks, without quantifying the

Figure 2. Margins of Error in Estimates of Governance in Selected Countries, 2006
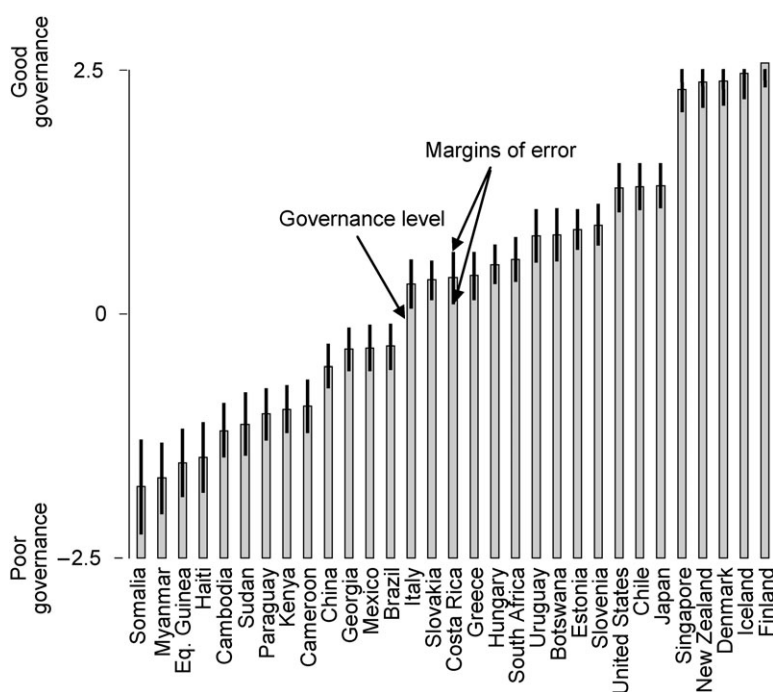
Source: Kaufmann, Kraay, and Mastruzzi. 2007a.

measurement error these rankings inevitably contain. This has contributed to a spurious sense of precision among users of these indicators and to an overemphasis on small differences across countries.

Of course, aggregate indicators have their own shortcomings. Foremost among them is the inevitable loss of specificity. Averaging one indicator of judicial corruption and another indicator of bureaucratic corruption arguably yields a more informative indicator of overall corruption, but not necessarily a more informative indicator of either of the two specific types of corruption. Averaging an indicator of freedom of the press with an indicator of electoral integrity yields a more informative indicator of overall democratic accountability, but not of either particular concept. For some purposes the broad aggregate indicators will be useful; for others the disaggregated indicators will be more useful. This is not a shortcoming, however, because virtually all aggregate governance indicators can readily be disaggregated into their constituent components, giving the user the freedom to choose the appropriate level of aggregation for the task at hand.[12]

The second concern with aggregate indicators is that their effectiveness at reducing measurement error depends crucially on the extent to which their underlying sources provide independent information on governance. Some types of expert assessments may make correlated errors in their governance rankings (although empirical evidence suggests that these error correlations are not likely to be very large). Aggregate indicators can mitigate only the component of measurement error that is truly independent across the different underlying indicators. This point is particularly relevant when contrasting multiple- and single-source aggregate indicators.

The WGI are multiple-source aggregate indicators, combining information from a large number of sources. In contrast, many other data sources report aggregates of their own subcomponents. For example, there is an aggregate CPIA rating in conjunction with the 16 underlying components, and there are six aggregate Global Integrity indicators, which combine information from more than 200 underlying individual indicators. All of the underlying individual indicators for a given country are scored by the same respondents. As a result, any respondent-specific biases are likely to be reflected in all of the individual indicators; the gain in precision from relying on the aggregate indicators from these sources will not be as large as when aggregate indicators are based on multiple sources.

In summary, aggregate governance indicators can play a useful role in synthesizing and summarizing the large variety of individual governance indicators. Using aggregate indicators is one way to exploit the complementarities between the different types of indicators (rules or outcomes, surveys or experts). Aggregation can also increase the precision with which these aggregate indicators measure broad but unobservable concepts of governance. Of course, for some purposes, more specific indicators are useful. It is thus important to be able to easily disaggregate aggregate indicators into their constituent components, as is the case with the WGI.

Moving Forward

A sobering picture emerges from this review: while most indicators of governance have many virtues, all face distinct challenges. Researchers, therefore, need to look at a variety of indicators and sources when monitoring or assessing governance across countries, within a country, or over time. A few principles may be useful as this work, as the use of governance indicators in public sector policy-making and civil society monitoring, continues.

Avoid false dichotomies. Too often, discussions of governance indicators overemphasize distinctions between types of governance indicators, with insufficient regard for the strong complementarities between them. Artificially, sharp

distinctions are often drawn between "subjective" and "objective" indicators of governance, when, in fact, virtually all indicators of governance rely on the judgments or perceptions of respondents in one way or another. In some cases, even the terminology is misleading. The recently released Ibrahim Index of African Governance, for example, touts itself as providing objective assessments of governance, even though its core governance components are based primarily on purely subjective data, including the Transparency International CPI and subjective ratings by the Heritage Foundation and the Economist Intelligence Unit.

Distinctions between aggregated and disaggregated indicators often have an artificial element also. Some aggregate indicators transparently disclose each disaggregated source, enabling users to take advantage of the complementarities between the two types of indicators and blurring the distinction between the two. For some purposes, it is useful to combine information from many individual indicators into some kind of summary statistics, while, for other purposes, the disaggregated data are of primary interest. Even where disaggregated data are of primary interest, however, it is important to rely on a number of independent sources for validation, because the margins of error and the likelihood of extreme outliers are significantly higher for a disaggregated indicator.

An excessively narrow emphasis on "actionable" indicators detailing specific policy interventions immediately under the control of governments can divert attention from equally important discussions of which of these indicators are "action worthy," in the sense of having significant impacts on outcomes of interest. The answer is often context-specific and rarely obvious *a priori*. Focusing too much on "actionable" indicators while downplaying scrutiny of outcome indicators may result in undue emphasis on measures that may not translate into concrete progress.

Use indicators appropriate for the task at hand. As with all tools, different types of indicators are suited for different purposes. Governance indicators can be used for regular cross-country comparisons. While many of these indicators have become increasingly specific, they often remain blunt tools for monitoring governance and studying the causes and consequences of good governance at the country level. For these purposes, a wide variety of innovative tools and methods of analysis has been deployed in many countries (reviewing these methods is beyond the scope of this survey). Examples of in-country tools include the World Bank's Investment Climate Assessments, the World Bank Institute's Governance and Anti-Corruption diagnostics, the corruption surveys conducted by some chapters of Transparency International, and the institutional scorecard carried out by the Public Affairs Center in Bangalore, India. Many project-specific interventions and diagnostics are possible to measure governance at this level.[13]

Public and professional scrutiny is essential for the credibility of governance indicators. Virtually all of the governance indicators listed in table 2 are publicly

available, either commercially or at no cost to users. This transparent feature is central to their credibility for monitoring governance. Open availability permits broad scrutiny and public debate about the content and methodology of indicators and their implications for individual countries. Many indicators are also produced by nongovernmental actors, making it more likely that they are immune from either the perception or the reality of self-interested manipulation on the part of the government. Scholarly peer review can also strengthen the quality and credibility of governance indicators. For example, articles describing the methodology of the Doing Business indicators, the Database of Political Institutions, and the WGI have appeared in peer-reviewed professional journals. Transparency with respect to details of methodology and its limitation is also essential for credible use of governance indicators. It is important that users of governance indicators understand fully the characteristics of the indicators they are using, including any methodological changes over time and time lags between the collection of data and publication.

It is thus of concern that some proposed and existing indicators of governance are insufficiently open to public scrutiny. While the recent disclosure of the World Bank's CPIA ratings for low-income countries represents a positive step, these indicators are being disclosed for only about half of the roughly 130 countries for which they are prepared each year, and none of the historical data from 2005 or earlier are publicly available. Historical data on the CPIA ratings of the African Development Bank and Asian Development Bank have also not been disclosed publicly. This is unfortunate, given that the decision to selectively disclose recent CPIA data and not to disclose historical CPIA data is made by the executive boards of these organizations and therefore reflects the desire of the very governments these ratings are supposed to assess. Regarding transparency, it is also of concern that although the Public Expenditure and Financial Accountability initiative has been ongoing since 2000, it had resulted in indicators and reports on just 42 countries as of March 2007, for only one period per country, only nine of them publicly available. Moreover, because these reports are prepared in collaboration with the governments in question, their credibility may not be the same as those associated with third-party indicators. Similar concerns affect recent Organization for Economic Co-operation and Development-led efforts to construct indicators of public procurement practices.

Transparently acknowledge margins of error of all governance indicators. All governance indicators include measurement error and so should be thought of as imperfect proxies for the fundamentals of good governance. This is not just an abstract statistical point, but rather one of fundamental importance for all users of governance indicators. Wherever possible, such margins of error should be explicitly acknowledged, as they are in the WGI, and taken seriously when the indicators are used to monitor progress on governance. At times the lack of disclosure of

margins of error is rationalized by suggesting that they would be missed by most readers. Experience with the WGI suggests that this is not the case, with many users recognizing and benefiting from this additional degree of transparency about data limitations.

Exploit the wealth of available indicators, recognizing that progress in developing new indicators is likely to be incremental. Much more work needs to be done to exploit the large body of disaggregated measures of governance already in existence. Linking disaggregated indicators to disaggregated outcomes, both across countries and over time, is likely to be an important area of research over the next several years that is likely to have important implications for policymakers.

There is also scope for developing new and better indicators of governance. Work to improve such indicators will be important, as indicators are increasingly used to monitor the success and failure of governance reform efforts. But given the many challenges of measuring governance, it is important to recognize that progress in this area over the next several years is likely to be incremental rather than fundamental. Alongside efforts to develop new indicators, there is also a case to improve existing indicators, particularly in increasing the periodicity of heretofore one-off efforts and in broadening their country coverage (covering industrial and developing economies), as well as covering issues for which data are still scarce, such as money laundering.

Notes

1. For surveys of and user guides to governance indicators, see UNDP (2005), Arndt and Oman (2006), and Knack (2006). Because of space constraints, no attempt is made here to review the important body of work focused on in-depth within-country diagnostic measures of governance that are not designed for cross-country replicability and comparisons.

2. A fuller compilation of governance datasets is available at www.worldbank.org/wbi/governance/data.

3. Indeed, this is reflected in the terminology of "actionable" governance indicators emphasized in the World Bank's Global Monitoring Report (World Bank 2006).

4. See King and Wand (2007) for a description of how this problem can be mitigated by the use of "anchoring vignettes" that provide a common frame of reference to respondents in interpreting the response scale. The basic idea is to provide an understandable anecdote or vignette describing the situation faced by a hypothetical respondent to the survey. For example, "Miguel frequently finds that his applications to renew a business license are rejected or delayed unless they are accompanied by an additional payment of 1,000 pesos beyond the stated license fee." Respondents are then asked to assess how great corruption as an obstacle is for Miguel's business, using a 10-point scale. Since all respondents use the scale to assess the same situation, this rating can be used to "anchor" their

responses to questions referring to their own situation.

5. These two indicators are measured as the average of 14 "in law" components and the 20 "in practice" components of the elections indicator of Global Integrity.

6. Starting with the 2005 data, both the African Development Bank and the World Bank have made their CPIA scores public. The African Development Bank does so for all borrowing countries; the World Bank does so only for countries eligible for its most concessional lending.

7. Kaufmann, Kraay, and Zoido-Lobatón (1999a) show how the estimated margins of error of their aggregate governance indicators would increase if they assume that the error terms made by individual data sources were correlated. Recently, Svensson (2005), Arndt and Oman (2006), and Knack (2006) have raised this criticism again, largely without the benefit of systematic evidence. Kaufmann, Kraay, and Mastruzzi (2007b) provide a detailed response.

8. This is not to say that all of the surveys used to measure governance are necessarily representative in any strict sense of the term. In fact, one general critique is that several large cross-country surveys of firms that provide data on governance are not very clear about their sample frame and sampling methodology. The Executive Opinion Survey of the World Economic Forum, for example, states that it seeks to ensure that the sample of respondents is representative of the sectoral and size distribution of firms (World Economic Forum 2006). But it reports that it "carefully select[s] companies whose size and scope of activities guarantee that their executives benefit from international exposure" (p. 133). It is not clear from their documentation how these two conflicting objectives are reconciled.

9. A simple example is that respondents are asked whether they have ever offered a bribe. But before answering, the respondent is instructed to privately toss a coin and to answer "yes" if either they have in fact offered a bribe, or the coin comes up heads. See Azfar and Murrell (2006) for an assessment of the extent to which randomized response methods correct for respondent reticence and an innovative approach to using this methodology to weed out less than candid respondents.

10. The assumption of a common error variance is necessary in this simple example with two indicators in order to achieve identification. In this example, just one sample correlation in the data can be used to infer the variance of measurement error; just one measurement error variance can thus be identified. In more general applications of the unobserved components model, such as the WGI, this restriction is not required because there are three or more data sources.

11. For details on this calculation, see Kaufmann, Kraay, and Mastruzzi (2004, 2006). Gelb, Ngo, and Ye (2004) perform a similar calculation comparing the African Development Bank and World Bank CPIA scores. Their conclusion that the CPIA ratings have little measurement error is driven largely by the fact that the authors focus on the aggregate CPIA scores, which are very highly correlated between the two institutions. The focus here is on one of 16 specific questions; at this level of disaggregation, the correlation between the two sets of ratings is considerably lower.

12. For example, virtually all of the individual indicators underlying the aggregate WGI are available at www.govindicators.org.

13. One of the best-known and best-executed recent studies of this type is a study of corruption in a local road-building project by Olken (2007).

References

Acemoglu, Daron. 2006. "Constitutions, Politics, and Economics: A Review Essay on Persson and Tabellini's *The Economic Effects of Constitutions*." *Journal of Economic Literature* 63(4):1025–48.

Arndt, Christiane, and Oman Charles. 2006. "Uses and Abuses of Governance Indicators." OECD Development Center Study, Organisation for Economic Co-operation and Development, Paris.

Azfar, Omar, and Peter Murrell. 2006. "Identifying Reticent Respondents: Assessing the Quality of Survey Data on Corruption and Values." University of Maryland, Department of Economics, College Park Maryland.

Gelb, Alan, Brian Ngo, and Xiao Ye. 2004. "Implementing Performance-Based Aid in Africa: The Country Policy and Institutional Assessment." World Bank Africa Region Working Paper 77, Washington, D.C.

Hellman, Joel, and Daniel Kaufmann. 2004. "The Inequality of Influence." In J. Kornai and S. Rose-Ackerman, eds., *Building a Trustworthy State in Post-Socialist Transition*. New York: Palgrave McMillan.

Kaufmann, Daniel, Aart Kraay, and Pablo Zoido-Lobatón. 1999a. "Aggregating Governance Indicators." Policy Research Working Paper 2195. World Bank, Washington, D.C.

———. 1999b. "Governance Matters." Policy Research Working Paper 2196. World Bank, Washington, D.C.

Kaufmann, Daniel, Aart Kraay, and Massimo Mastruzzi. 2004. "Governance Matters III: Governance Indicators for 1996, 1998, 2000 and 2002" *World Bank Economic Review* 18(2):253–87.

———. 2005. "Governance Matters IV: Governance Indicators for 1996–2004." Policy Research Working Paper 3630. World Bank, Washington, D.C.

———. 2006. "Governance Matters V: Governance Indicators for 1996–2005." Policy Research Working Paper 4012. World Bank, Washington, D.C.

———. 2007a. "Governance Matters VI: Aggregate and Individual Governance Indicators for 1996–2006." Policy Research Working Paper 4280. World Bank, Washington, D.C.

———. 2007b. "The Worldwide Governance Indicators Project: Answering the Critics." Policy Research Working Paper 4149. World Bank, Washington, D.C.

Kautilya. 1992. [400 B.C.E.] *The Arthashastra*. New Delhi, India: Penguin Classic Edition.

King, Gary, and Jonathan Wand. 2007. "Comparing Incomparable Survey Responses: Evaluating and Selecting Anchoring Vignettes." *Political Analysis* 15(1): 46–66.

Knack, Steven. 2006. "Measuring Corruption in Eastern Europe and Central Asia: A Critique of the Cross-Country Indicators." Policy Research Department Working Paper 3968. World Bank, Washington, D.C.

North, Douglass. 2000. "Poverty in the Midst of Plenty." *Hoover Institution Daily Report*, October 2. (www.hoover.org.)

———. 2007. "Monitoring Corruption: Evidence from a Field Experiment in Indonesia." *Journal of Political Economy* 115(2): 200–49.

Persson, Torsten, and Guido Tabellini. 2005. *The Economic Effects of Constitutions*. Cambridge, Mass.: MIT Press.

Razafindrakoto, Mireille, and François Roubaud. 2006. "Are International Databases on Corruption Reliable? A Comparison of Expert Opinion Surveys and Household Surveys in Sub–Saharan Africa." Development Research Institute, Development Institutions and Long-Term Analysis (IRD/DIAL), Paris.

Svensson, Jakob. 2005. "Eight Questions about Corruption." *Journal of Economic Perspectives* 19(3): 19–42.

UNDP (United Nations Development Programme). 2005. *Governance Indicators: A Users Guide*. New York: UNDP.

World Bank. 1992. *Governance and Development*. Washington, D.C.

———. 2002. *Building Institutions for Markets*. New York: Oxford University.

———. 2006. *Global Monitoring Report*. Washington, D.C.

————. 2007. "Strengthening World Bank Group Engagement on Governance and Anticorruption." Joint Ministerial Committee of the Boards of Governors of the Bank and the Fund on the Transfer of Real Resources to Developing Countries, Washington, D.C. [www.world-bank.org/html/extdr/comments/governancefeedback/gacpaper.pdf].

World Economic Forum. 2006. *The Global Competitiveness Report 2006–2007*. New York: Palgrave Macmillan.