

# What Is Governance and How Do We Measure It?

*Outline of Presentation at Workshop on Governance and Development  
Dhaka, November 11-12, 2006*

Aart Kraay, The World Bank<sup>1</sup>

## 1. What Is Governance?

### 1.1 Definitions

Although the terms "governance" and "institutional quality" are widely used in discussions of economic development, they lack widely-accepted definitions. Frequently-used definitions of the concepts tend to be quite broad. For example, Douglass North implicitly defines institutions, together with a link to economic development, as follows:

*"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"*

In an ongoing project to construct cross-country measures of governance that I will discuss in more detail below, we used a similarly-broad definition of governance as

*"...the traditions and institutions by which authority in a country is exercised. This includes (1) the process by which governments are selected, monitored and replaced, (2) the capacity of the government to effectively formulate and implement sound policies, and (3) the respect of citizens and the state for the institutions that govern economic and social interactions among them."*

Despite these rather general definitions, there is considerable agreement about certain broad features of what constitutes "good governance". For example, few would argue that a rapacious state which fails to protect property rights or abide by the rule of law is an example of good government, and most people would agree that some form of accountability of government to citizens is a feature of good governance. But at the same time there is scope for considerable disagreement over the precise set of formal and informal rules that bring about a respect of property rights or an accountable government.

### 1.2 From Institutions to Policies

---

<sup>1</sup> 1818 H Street N.W., Washington, DC 20433 U.S.A., akraay@worldbank.org. This presentation draws on a series of papers coauthored with Daniel Kaufmann and Massimo Mastruzzi, that are listed in the references. The views expressed here are the author's, and do not reflect those of the World Bank, its Executive Directors, or the countries they represent.

When thinking about how to define and measure governance or institutional quality, it can be helpful to think about these concepts as organized along a continuum. At the one extreme, there are very "deep" and concrete institutional features of a country, such as the features of its constitution or legal system. Does the constitution provide for a presidential or a parliamentary system? Does the electoral system provide for proportional representation or "first-past-the-post" representation? Is the central bank formally independent from other branches of government? Does the country have a common-law system or does it follow a Napoleonic code? These features are readily-observed and identifiable as institutions, and there is a rich body of empirical work relating cross-country differences in these types of institutions to various outcomes of interest.

At the other extreme are policy outcomes, such as the size of the budget deficit or the rate of inflation, or even the level of development itself. Few would argue that such outcomes themselves should be thought of as "governance". And if we define governance too much in terms of outcomes, then the links between governance and outcomes risks becoming tautological.

But in between these extremes lie a huge variety of possible things that one could measure and identify as governance. Take for example features of the regulatory environment, such as the number of procedures required to register a business. On the one hand, by virtue of being a set of rules governing economic interactions, one could think of this as an institution or governance. On the other hand, the regulatory environment is the outcome of a political process, and the way in which the formal and informal institutions of governance manage to resolve the tension between competing interests will affect the type of regulatory environment we observe. So in this sense the regulatory environment is more like and outcome of deeper institutions.

We are unlikely in this conference to be able to resolve such definitional ambiguities. While they are interesting, they should not prevent us from thinking about what features of governance we can usefully measure, and with which kinds of data. Proceeding in this way without clear prior definitions carries the risk that governance will end up being defined implicitly by the measures that we construct. But this may not be such a bad thing if the measures we construct are interesting, linked to development outcomes that we care about, and potentially amenable to reforms and improvement.

## **2. Measuring Governance**

Measuring governance is important, given the demonstrated strong links between good governance and good development outcomes. Steve Knack will discuss the relationship between governance and growth in more detail later in the conference. For now, I simply note that a wide array of measures of governance tend to be better in rich countries, and the balance of a large body of scholarly research suggests that much of the causality in this relationship runs from governance to development.

## 2.1 The Worldwide Governance Indicators Project

A dozen years ago few internationally comparable measures of governance or corruption existed. Since then, there has been an explosion of empirical research aimed at measuring governance and monitoring country progress. The Worldwide Governance Indicators (WGI) are one contribution to this work. They are based on an ongoing research program of the World Bank Institute and the Research Department of the World Bank, that I initiated in the late 1990s with Daniel Kaufmann, and with the assistance of Pablo Zoido-Lobaton and Massimo Mastruzzi. Sometimes referred to as the “KK,” “KKZ,” or “KKM” indicators after the authors' initials, they have been compiled since 1996 and measure the quality of governance in over 200 countries, based on more than 30 data sources produced by 25 different organizations worldwide.

The WGI capture six dimensions of governance:

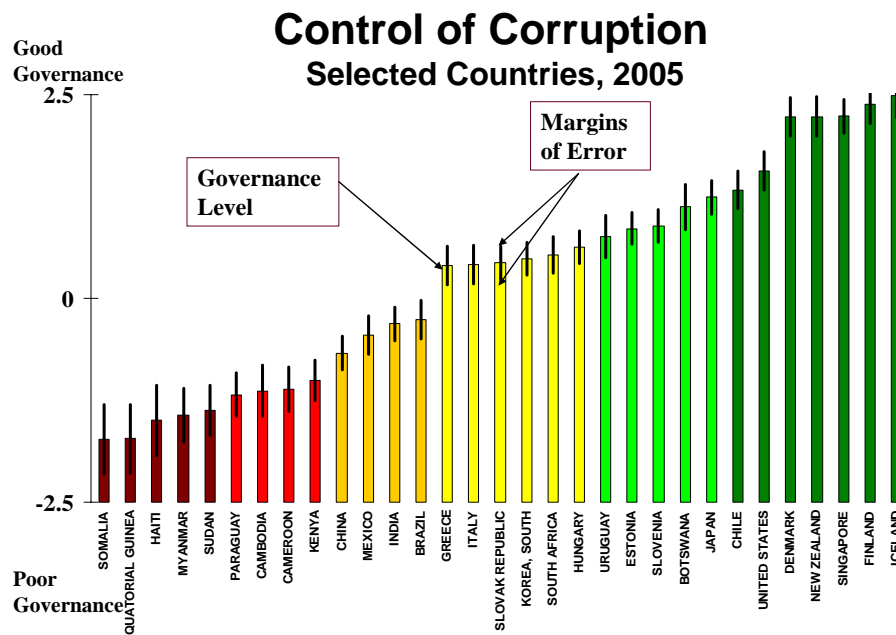
- *Voice and accountability*, the extent to which a country's citizens are able to participate in selecting their government, as well as freedom of expression, freedom of association, and a free media
- *Political stability and absence of violence*, perceptions of the likelihood that the government will be destabilized or overthrown by unconstitutional or violent means, including domestic violence and terrorism
- *Government effectiveness*, the quality of public services, the quality of the civil service and the degree of its independence from political pressures, the quality of policy formulation and implementation, and the credibility of the government's commitment to such policies
- *Regulatory quality*, the ability of the government to formulate and implement sound policies and regulations that permit and promote private sector development
- *Rule of law*, the extent to which agents have confidence in and abide by the rules of society, and in particular the quality of contract enforcement, the police, and the courts, as well as the likelihood of crime and violence
- *Control of corruption*, the extent to which public power is exercised for private gain, including both petty and grand forms of corruption, as well as “capture” of the state by elites and private interests

The WGI have been published every other year between 1996 and 2002, and annually for 2002-2005. The latest update for 2005 was released in September of 2006. With this latest release, for the first time virtually all of the individual data sources underlying the aggregate indicators are, along with the aggregate indicators, available electronically at [www.govindicators.org](http://www.govindicators.org). This large collection of detailed individual and aggregate indicators covering a decade is one of the world's largest publicly available compilations of data on governance.

The aggregate indicators are based on several hundred underlying variables that reflect perceptions of a wide range of governance issues. For the 2006 release, which include data through the end of 2005, the governance indicators are drawn from 31 separate data sets maintained by 25 different organizations worldwide. The data consist of surveys of firms and individuals, as well as the assessments of commercial risk-rating agencies, nongovernmental organizations and think tanks, and multilateral aid agencies. Almost all sources are available annually, and the WGI draw on only the most recent data available.

Each of the many individual indicators becomes part of one of the six aggregate indicators, on the premise that any single indicator provides only a partial measure of the broader notion of governance to which it is assigned. For example, we have several different cross-country surveys of firms and individuals that ask respondents about their perceptions of and/or experiences with corruption. We also have several commercial risk rating agencies that provide "expert assessments" of corruption based on a network of information sources. Our view in constructing the WGI is that each one of these measures provides a useful, but imperfect, signal of corruption, and that by averaging these measures together we can obtain more reliable cross-country comparisons of corruption.

We use a statistical methodology known as the "unobserved components model" to construct aggregate indicators from the individual measures. The aggregate indicators are weighted averages of the underlying data, with weights reflecting the precision of the individual data sources. Aggregate indicators are more informative about the six broad dimensions of governance than any individual measure can be, since they average information from many different sources.



Source for data: 'Governance Matters V: Governance Indicators for 1996-2005', D. Kaufmann, A. Kraay and M. Mastruzzi, September 2006, forthcoming.  
 Note: Colors are assigned according to the following criteria: Dark Red, below 10<sup>th</sup> percentile rank among all countries in the world; Light Red between 10<sup>th</sup> and 25<sup>th</sup>; Orange, between 25<sup>th</sup> and 50<sup>th</sup>; Yellow, between 50<sup>th</sup> and 75<sup>th</sup>; Light Green between 75<sup>th</sup> and 90<sup>th</sup>; Dark Green above 90<sup>th</sup>.

The unobserved components model also generates margins of error for the estimates of governance for each country. The margins allow users to identify not just a single rating for each country, but also a range of statistically likely ratings, as illustrated in the accompanying figure on corruption for 2005. The vertical axis plots the point estimate for each country, together with the statistically likely range of values as a vertical black line. There is a 90 percent chance that the true value of governance for the country lies within that range.

The ranges are not trivial, which means that some of the smaller differences in governance across countries shown in the figure are not statistically meaningful. For example, the likely range of values for corruption for Sweden, Denmark, Singapore, and Finland all overlap substantially, suggesting that the slight differences in their country rankings should not be overinterpreted as capturing important differences. But in many other comparisons of countries, these likely ranges of values do not overlap, indicating much more important differences across countries. For example, differences between Sweden and Japan, or Japan and South Africa, or South Africa and Brazil, or Brazil and Kenya, are all much more significant. The majority of country comparisons are like this. The 2005 control of corruption indicator, for example, which covers 204 countries, potentially allows for 20,706 pairwise comparisons between pairs of countries. The differences revealed in roughly three-quarters of those comparisons are statistically important at the 75 percent confidence level.

Margins of error are not unique to the Worldwide Governance Indicators, nor are they due to the use of perceptions data. Instead, they reflect the pervasive difficulties of measuring governance itself. No single measure can perfectly capture complex concepts such as corruption, government effectiveness, voice and accountability, or the investment climate.

With the WGI, we have been forthright in acknowledging the imprecision of our estimates of corruption and other dimensions of governance, and we encourage other raters to do the same. Unfortunately, most existing measures of governance (or investment climate) simply report country rankings without regard for the inevitable uncertainty associated with such rankings. The failure to report margins of error encourages a sense of false precision, which permits excessive emphasis on marginal differences in country rankings that may in fact be due to measurement error.

## **2.2 Other Measures of Governance**

I have devoted most of this presentation to discussing the Worldwide Governance Indicators because they are the ones that I am most familiar with. But I do not want to leave the impression that these are the only types of governance indicators available. Rather, the WGI are part of a very large set of different possible measures of governance, and I cannot do justice to all of them here. Fortunately, there are several useful recent summaries of different types of available governance data, including compendiums by the OECD (Arndt and Oman (2006), "Uses and Abuses of Governance Indicators"), the EC and UNDP (2004, "Governance Indicators: A Users' Guide"), and the World Bank (2006 "Global Monitoring Report"). Here I want to simply point out in very broad terms

the different types of governance data that are available, with a view to discussing some of their relative strengths and weaknesses in the next section.

- *Features of Political Systems* There are several such databases, notably one compiled by colleagues at the World Bank (Keefer et. al. "The Database of Political Institutions"). These measure things like details of constitutions, voting procedures, the presence of checks and balances in the political system, etc. Thinking about the measurement continuum discussed above, this kind of data is closer to the "deep institutions" end.
- *Features of Legal and Regulatory Systems.* The most famous of these is the World Bank's "Doing Business" database, which seeks to measure in a consistent way the "de jure" legal and regulatory environment in different areas, such as procedures to start a business, rules governing the hiring and firing of workers, the function of the courts as a dispute settlement mechanism.
- *Investment Climate Surveys.* These surveys of firms have been carried out by the World Bank in over 80 developing countries over the past several years, using a common core questionnaire that allows cross-country comparisons of firm responses. These surveys are not focused exclusively on governance, but many of the questions firms are asked are informative about dimensions of governance, including the regulatory environment, property rights protection, and corruption.
- *Governance Diagnostic Surveys* These surveys have been carried out in roughly 20 developing countries. Unlike these other sources they are not designed to be internationally comparable, but rather to give detailed country-specific information on perceptions of governance, focusing on responses of firms, individuals, and public officials, that can be used to design country-specific governance reform strategies.

In addition, a lot of recent research tries to measure different aspects of governance, and especially corruption, in innovative direct ways. These are typically one-off efforts to measure corruption in a very specific environment, and because of their specificity they do not lend themselves to cross-country comparisons of corruption. Rather they can be useful in identifying specific factors driving corruption in specific circumstances.:

- *Comparing Inputs and Outputs.* Golden and Picci (2005) comparing recorded spending on infrastructure across regions in Italy with direct measures of the stock of infrastructure. They find that gaps between past spending and actual stocks are much higher in the south of Italy, which they interpret as evidence of greater corruption there. Olken (2005) conducted an innovative experiment using a decentralized local road-building project in Indonesia. Some villages were threatened with audits, while in other villages participation in meetings to oversee the road project was encouraged. Then the authors audited the individual road projects, comparing actual road quality with recorded expenditures, and found that threats of audit significantly reduced the discrepancy between the two.
- *Corruption in the United Nations* is the topic of two recent papers. One paper by Hsieh and Moretti (2005) documents evidence of corruption in the UN Oil-for-Food program in Iraq by looking at discrepancies between prices of Iraqi crude oil

and oil from other sources. Kuziemko and Werker (2006) document a strong correlation between rotation of temporary members on to and off of the UN Security Council and increases (and subsequent declines) in foreign aid receipts from the United States, which they interpret as evidence of "vote-buying" in the UN.

### 3. Issues in Measuring Governance

I want to conclude by discussing a number of issues that arise in measuring governance, focusing particularly on the distinction between subjective or perceptions-based measures of governance and more objective measures. In particular I want to address two broad questions: (i) are perceptions-based measures of governance reliable? and (ii) why don't we have more objective indicators of governance?

***Do subjective or perceptions-based measures of governance provide reliable information, or do they simply reflect vague and generic perceptions—or worse, simply hearsay and rumor?***

The Worldwide Governance Indicators are a compilation of information and perceptions from a very diverse group of respondents, collected in a large number of surveys and other cross-country assessments of governance. Some of these instruments capture the views of firms, individuals, and public officials in the countries being assessed. Others reflect the carefully considered views of NGOs and aid donors with considerable experience in the countries being assessed. Still others are based on the assessments of commercial risk-rating agencies, which face a "market test." If their information were worthless hearsay, it would find few buyers. The extent of agreement among these very diverse stakeholders is striking. For corruption, for example, the assessments of several commercial risk-rating agencies are correlated at 80 percent or higher with a totally independent survey of firms across countries.

Perceptions data are very often specific and experiential, rather than vague and general, and many specific questions go into each aggregate indicator. For example, the Global Competitiveness Survey coordinated by the World Economic Forum asks questions like this: "*When firms like yours do business with the government, how much of the contract value must they offer in additional payments to secure the contract?*" Equally specific questions are posed by other firm surveys like the World Bank's Business Environment and Enterprise Performance Survey (BEEPS). Household surveys such as the Gallup Organization's Voice of the People and Global Barometer Surveys and Latinobarometro ask respondents to report percentages of corrupt officials or a specific number of times they witnessed acts of corruption.

Perceptions themselves are worth measuring. If voters do not believe that elections will reflect the will of the people, they are unlikely to bother to vote. If citizens believe the courts and police to be corrupt and inefficient, they will be reluctant to use their services. If entrepreneurs believe that business registration procedures and business regulation are costly and inefficient, they may choose to remain in the informal sector rather than comply with such rules.

Another concern is that perceptions data from commercial sources or NGOs can reflect the ideological biases of the institutions providing the data. For example, commercial data sources that provide information to businesses, and business-oriented NGOs or think-tanks, are often accused of having ideological biases favouring right-wing governments, and that this taints their ratings. We have looked at this question in detail using the data underlying the WGI. Specifically we asked whether such data sources give significantly higher ratings to right-wing governments, after controlling for "neutral" assessments provided by cross-country surveys. We found very little evidence of such biases.

A related concern is that commercial data sources and such NGOs engage in a form of "group-think" whereby their assessments are based more on each others' assessments than on the actual situation in the countries they are assessing. One way to test for this is to see whether commercial risk rating agencies' assessments of governance are more correlated with each other than they are with more "neutral" data coming from cross-country surveys of firms or individuals. It turns out that this is not the case -- typically the commercial risk rating agencies are about as correlated with each other as they are with third-party more "neutral" sources.

***Shouldn't we try to measure governance with "hard" objective data rather than "soft" perceptions data?***

For many dimensions of governance, relevant objective data are difficult to obtain. Consider the difficulty of measuring corruption objectively. Almost by definition, corruption leaves no paper trail and so is hard to measure objectively. As discussed above, there are innovative approaches to detecting the tell-tale signs of corruption in specific contexts. But these approaches are very difficult to scale up to enable cross-country comparisons for a large set of countries.

Even where objective measures *are* available, they provide only imperfect proxies for real conditions on the ground (of course the same is true for perceptions-based data which has potential problems of its own). For example, the constitutional limits on executive authority in a country, the laws governing judicial independence, or the regulations governing business entry may correspond very poorly with the actual application of those rules and procedures.

As a specific example, we have studied the relationship between formal rules governing business entry in developing countries and surveys of firms' perceptions of the ease or difficulty of business entry. Among industrial countries, we found that the correlation between firms' perceptions of their "on the ground" experience with business entry regulations corresponded quite closely with the "on the books" formal rules. But in developing countries, the correlation is quite low, capturing significance discrepancies between what appears on the books and what is practiced on the ground. Much of the difference can be traced to corruption, which leads to a subversion of official rules. There should be no presumption, therefore, that objective data are necessarily more informative than reports from experts, citizens, or firms with real-world experience. The Worldwide

Governance Indicators reflect this wealth of experience by relying on such perceptions data.

## References

- Hsieh, Chiang-Tai and Enrico Moretti (2006). "Did Iraq Cheat the United Nations? Underpricing, Bribes, and the Oil for Food Program". *Quarterly Journal of Economics*, forthcoming.
- Golden, Mirian and Lucio Picci (2005). "Proposal for a New Measure of Corruption, Illustrated with Italian Data". *Economics and Politics*. 17(1):37-75
- Kaufmann, Daniel, Aart Kraay, and Massimo Mastruzzi. 2004. "Governance Matters III: Governance Indicators for 1996, 1998, 2000, and 2002." *World Bank Economic Review* 18: 253–287.
- . 2005. "Governance Matters IV: Governance Indicators for 1996–2004." World Bank Policy Research Working Paper 3630. Washington, DC.
- . 2006. "Governance Matters V: Governance Indicators for 1996–2005." World Bank Policy Research Working Paper, September. Washington, DC.
- . 2006. "Measuring Governance Using Perceptions Data." In Susan Rose-Ackerman (ed.), *Handbook of Economic Corruption*. Edward Elgar.
- . 2006. "Measuring Corruption: Myths vs. Realities." *Development Outreach* (World Bank), September.
- Olken, Ben (2005) "Monitoring Corruption: Evidence from a Field Experiment in Indonesia". NBER Working Paper No 11753.