

GOVERNANCE

BASEL INSTITUTE ON GOVERNANCE

Conference «Non-State Actors as Standard Setters: The Erosion of the Public-Private Divide» | February 8-9 2007 | Hotel Hilton Basel, Switzerland

On measuring governance. Framing issues for Debate

Daniel Kaufmann and Aart Kraay, World Bank Institute

Issues paper for January 11th, 2007 Roundtable on Measuring Governance, Hosted by the World Bank Institute and the Development Economics Vice-Presidency of The World Bank

The World Bank, 1818 H Street NW, Washington, DC 20433 dkaufmann@worldbank.org, akraay@worldbank.org. The views expressed here are the authors, and do not reflect those of the WorldBank, its Executive Directors, or the countries they represent.

Our work on developing governance indicators has made us acutely aware of the many difficulties that arise in efforts to measure governance, and in using governance indicators to inform policy decisions.¹ In this short issues paper we offer three simple principles for producers and users of governance indicators that may both provide a useful summary framing these main difficulties, and at the same time suggest ways forward in improving governance indicators. Our major theme is that since all governance indicators are limited in various ways, it is important to recognize and exploit the complementarities between alternative approaches to measuring governance.

1. Measurement error is pervasive in all efforts to measure governance and investment climate.

There exists by now a very large variety of different cross-country and within country measures of governance. Among them, the Worldwide Governance Indicators project that we began in the late 1990s (WGI) is nearly unique in that it explicitly recognizes the unavoidable imprecision that is involved with measuring governance.² Yet producers and users of governance data should not misinterpret the absence of explicitly-reported margins of error in other indicators as evidence

of precision or accuracy. Rather, it is important to keep in mind that *all* measures of governance will ultimately be imperfect proxies for the dimensions of governance one really wants to measure. After all, governance writ large is a concept that defies easy definition, and even some of the commonly accepted dimensions of governance, such as democratic accountability, government effectiveness, or rule of law, are themselves subject to definitional ambiguities. Even in the case of corruption, which does have an accepted definition as the use of public office for private gain, there is latitude in interpreting such a definition. Furthermore, one wants to distinguish between different types of corruption (e.g. petty vs. judicial vs. "grand" corruption, state capture), and it should not be expected that an indicator of *one* type of corruption will be a good proxy for *other* types.

These definitional challenges should not lead to paralysis in measurement, since many manifestations clearly related to governance can be measured. Rather, they indicate that some realism is in order about the extent to which these broad notions of governance can be accurately measured. In particular, given that governance is a complex and multi-dimensional concept, it is unrealistic to expect that any single indicator of governance can provide a precise all-encompassing measure. Rather, any governance indicator will be subject to two key sources of imprecision, i.e., to two types of measurement error, namely:

First, any specific proxy of governance will itself have measurement error relative to the specific concept it seeks to measure, due to intrinsic measurement challenges. For example, a survey question about corruption will have the usual sampling error associated with it. Similarly, efforts to objectively document the specifics of the institutional environment or regulatory regime, such as business entry regulations to take just one example, will face challenges in coming up with a factually accurate description of the relevant laws and regulations in each setting. Or for instance measures of the composition and volatility of public spending, which are sometimes interpreted as indicators of undesirable policy instability, are subject to all of the usual difficulties in measuring public spending (and also the varying degrees of inaccuracies in official statistics) consistently across countries and over time. And finally there can simply be differences of opinion between respondents --

¹ For details on the Worldwide Governance Indicators (WGI) project visit www.govindicators.org. Through such website, in addition to having access to all the WGI data and papers (including specifically responding to critics), one can also access to relevant papers and reports by others, including the GMR.

² The only other exception we are aware of is TI's Corruption Perceptions Index (CPI), which now does report a measure of the dispersion across individual data sources in their country ratings.

GOVERNANCE

BASEL INSTITUTE ON GOVERNANCE

Page 2

for example different groups of experts might come up with very different assessments of the same phenomenon in a particular country.³

Second, any proxy for governance is by definition an imperfect measure of broader concepts of governance. A specific survey question about corruption in public procurement will not be fully informative about overall corruption in the broader public sphere. Information about the statutory requirements for business entry regulation need not reflect the actual practice of how these requirements are implemented on the ground, nor are they informative about regulatory burdens in other areas.

These sources of measurement error are neither surprising, nor should they paralyze efforts to measure governance—all of which necessarily are imperfect proxies for broader and fundamentally unobservable concepts. But it means that one should fully account for the margins of error for all types of governance indicators. In the context of policy discussions this also points to the importance of avoiding over-interpretation of relatively small changes in governance proxies over time (or across countries), as they may not signify meaningful changes in underlying governance structures.

Indeed, we submit that one of the virtues of the WGI is its emphasis on margins of error and their interpretation when assessing the significance of cross-country and over-time comparisons of governance. In fact, in other work we have provided evidence suggesting that objective indicators have margins of error that are at least as large as those of subjective indicators. This universality of margins of error across individual indicators of governance and investment climate is usually ignored. The lack of disclosure of such margins of error often conveys a false sense of precision, and encourages over-interpretation of small (and likely statistically and practically insignificant) changes from one year to the next ('elevator economics'), and also of small differences across countries.

2. There are no silver bullets in measuring governance.

The many efforts to measure governance in the past 10 years have often featured strong claims on the part of

proposers, and providers, of new measures regarding their merits relative to existing indicators. An unfortunate consequence of these competing claims is that for many years it has created the expectation in some circles that there are - - or soon will appear -- "silver bullet" governance indicators that supersede and vastly improve on the current set of available governance indicators. For example, several years ago, not long after the launch of the WGI in the late nineties, there was a concerted push for so-called "second-generation" governance indicators that promised to supersede the WGI.

Similarly, more recently there has been a push for "actionable" governance indicators (more on this below). And there are continual debates over the relative merits of "subjective" versus "objective", and "aggregate" versus "individual" indicators. In sum, our sense is that too much of this discussion has tended to be of a confrontational nature, based on the mistaken premise that alternative governance indicators are necessarily substitutes for each other, and that eventually only one such indicator can be useful. By implication, too little attention has been paid to the complementarity among these indicators.

We are acutely aware of the limitations of various types of governance indicators, including the WGI. Innovative efforts to come up with alternative measures of governance, as well as improving upon what exists, ought to be encouraged and welcomed. At the same time, it is not helpful to set up such false dichotomies between alternative types of governance indicators. Rather, we emphasize that there are important complementarities between different governance and investment climate indicators, and also in recognizing that different indicators are either designed for different purposes, or are at least more appropriate than others for particular objectives. We illustrate this general point by pairing various types of indicators and providing several specific examples:

Aggregate versus individual indicators. Aggregate indicators that average various underlying individual indicators of governance have the advantage of (i) very broad country coverage, (ii) providing a useful summary of the myriad individual indicators, (iii) averaging out and so reducing measurement error and otherwise reducing the influence of idiosyncracies of individual data sources⁴, and, (iv) allowing for the calculation of explicit margins of error. At the same time, aggregate indicators have drawbacks, notably (i) the difficulty in interpreting

³ An example here is the high, but not perfect, correlation between individual components of the World Bank and African Development Bank's CPIA ratings across countries in Africa. A further example of differences of opinion comes from the latest Global Integrity Index's rating of the integrity of elections. The correlation across countries in Africa of this expert assessment with a survey question asking households whether elections are free and fair is nearly zero. In both examples, we do not suggest that the one source or the other has the monopoly on truth. Rather, both sources are measuring similar things, but with unavoidable margins of error.

⁴ As an example of such anomalies that tend to occur for some countries in individual indicators, consider the BEEPS survey of firms, where Belarus and Uzbekistan rank 4th and 6th *best* out of 27 countries on questions about corruption, while a commercial risk rating agency (DRI) lowly ranks these countries as 23rd and dead last, respectively. Since the latter ranking more closely reflects the consensus of our other data sources, these two countries score poorly in the WGI Control of Corruption indicator.

GOVERNANCE

BASEL INSTITUTE ON GOVERNANCE

Page 3

such summary statistics and their changes over time, and, (ii) the difficulty in understanding how reforms in specific areas will affect a country's ranking on aggregate indicators. We are aware of these difficulties, and in our work with the WGI we have made a variety of efforts to ease these drawbacks. Most notably with the latest release of WGI we have made publicly available virtually all of the underlying data sources on which we rely.

Aggregate and individual indicators: complementarities within a given set of indicators. It is apparent, however, that neither aggregate nor individual indicators are unambiguously a better tool for assessing governance or identifying reform priorities, and so there is no clear reason in general to prefer the one type over the other. In fact, some of the most useful aggregate indicators are ones that can themselves be easily disaggregated into their constituent parts. The Doing Business index, the WGI project, the Global Integrity Index, the Freedom House Freedom in the World ratings, and (for low-income countries) the World Bank's Country Policy and Institutional Assessment (CPIA) ratings are all examples of aggregate measures for which the underlying components are also publicly available. We do note however that a key distinction between the WGI and these other data sources is that the others draw their individual components from the same set of respondents. This of course increases the risk that respondent-specific biases creep into all of the individual indicators as well as the aggregate indicator. The WGI is in this sense unique in that it aggregates the views of many different responding organizations, and so is less likely to be affected by the particular biases of a single set of respondents.⁵

Subjective versus objective indicators. One of the most heated debates among users and producers of governance indicators is over the relative usefulness of subjective or perceptions-based measures of governance versus objective or fact-based measures. As producers of the WGI, which is an aggregate set indicators based on 'perceptionsbased' data, we view these two types of measures – subjective and objective -- as complementary rather than alternative approaches.⁶

⁵ TI's CPI, focused on one variable (corruption), also aggregates across different respondents, but unlike the WGI it does not make the data from their underlying data sources publicly available.

⁶ We also note that the distinction between "subjective" and "objective" indicators can be blurry. Some subjective measures nevertheless elicit very specific quantitative responses, for example survey questions that ask firms what fraction of a contract value is typically required as bribes. And notionally objective measures, such as an assessment of the steps a hypothetical firm might need to take in order to fire a hypothetical worker, inevitably involve judgment on the part of the legal expert filling the questionnaire. After all, lawyers generally do refer to their views as "opinions"!

To take a specific example, the fact-based documentation of the *de jure* state of business regulation across countries in the Doing Business project is very useful in identifying specific regulatory bottlenecks that policymakers can address in law. But at the same time it is important to recognize that there are important gaps between laws on the books and their implementation on the ground. In other papers we have for example documented that the correlation of firms' perceptions of the ease of business entry or the burden of taxation in developing countries are not very highly correlated with the corresponding statutory requirements, and that the discrepancy between the two can substantially be explained by the prevalence of informality or corruption which subverts the implementation of statutory rules.⁷

Further, we argue that perceptions matter in their own right, since after all firms and individuals take actions based on their perceptions. And finally, for some areas of governance such as corruption that leave no "paper trail", it is difficult to come up with alternatives to perceptions data. In short, neither subjective nor objective measures in isolation should be thought of as a silver bullet for the problem of measuring governance.

Timely cross-country comparisons versus detailed diagnostics. The simple point is that different types of governance indicators are appropriate for these two different purposes, without one type being unambiguously superior to the other. For some purposes, such as global benchmarking of countries on a regular basis, it is important to have governance indicators with broad cross-country coverage that are regularly updated in order to reflect changes for better or worse in the countries being assessed. Several of the aggregate indicators we have already discussed, such as the WGI, Doing Business, and the CPIA ratings can be used for these purposes. For other purposes we need very detailed country-specific assessments of governance challenges in a country, relying on tools such as governance diagnostic surveys, investment climate surveys (ICAs), public expenditure tracking systems (PETS), Governance Assessments, and others. And for project-specific monitoring purposes, initiatives such as the comparison of expenditures versus actual materials used in road building (in an Indonesia project) are very valuable.

Ideally a very detailed country assessment would be carried out annually in a very large number of countries, hopefully using a common methodology that would allow the same tool to be used both for within-country

⁷ Various papers are at www.govindicators.org More generally, note that often there is the common misconception that accuracy or precision can be equated with 'hard', official or objective (vs. subjective) indicators, or that precision can be identified with specificity (or individual indicators). Similarly, the absence of disclosed margins of error in a data source tends to signal a false sense of precision.

GOVERNANCE

BASEL INSTITUTE ON GOVERNANCE

Page 4

diagnostics, as well as cross-country comparisons and high-frequency monitoring. Realistically, however, this would be a very large, costly and complex endeavor which is unlikely to happen in the near future. For instance, the most frequent time series coverage of surveys in the ICA family are those in the transition economies, and even in this case the coverage is only once every three years, and following a format somewhat different than for the other ICA surveys. Further, while very useful for monitoring expenditure leakages in a particular sector such as health or education, the country coverage of detailed fiscal diagnostics such as PETS remains limited – and it is context specific (as also are project-specific monitoring tools).

While these tools will continue to be very valuable as ways to assess governance challenges in individual countries, they are not likely to provide the very regular and large country coverage that is needed for other purposes. Nor, symmetrically, are large annual cross-country indicators such as those discussed above, by themselves going to provide an in-depth country diagnostic or a sufficient basis to identify reform priorities in individual countries. Again, there are no silver bullets, and different tools are appropriate for different purposes. For in-depth governance assessments it is better to focus much more on the complementarities among the various tools and indicators – aggregate and disaggregate, subjective and objective.

3. The links from governance to development outcomes are complex.

This observation should come as no surprise to scholars or policymakers that have grappled with the difficulties of reforming governance. But it also has implications for how to go about measuring governance, both within and across countries. For example, in some circles the current mantra (having evolved from past efforts proposing ‘second generation’ indicators) is to produce “actionable” indicators of governance. This means indicators that measure specific things under the control of policymakers. Examples include the statutory rules governing the business environment in the Doing Business effort, measures of civil service recruitment and turnover practices, specifics of budget procedures, the OECD-DAC Procurement Indicators, the Public Expenditure and Financial Accountability (PEFA) indicators, and many others. Undoubtedly it is useful to compile accurate information on such indicators, both to inform policymakers in individual countries and to allow for cross-country comparisons.

However, an oft-ignored caveat is that simply because something can be measured does not mean that it is an important constraint on good governance. In short, not all “actionable” indicators need also be “action-worthy”. To illustrate, one can measure whether or not a country has an independent anticorruption commission, but we know

that this is no guarantee that in any particular country the creation of such a commission would help to reduce corruption. Alternatively, we can in principle measure the speed of judicial proceedings, but it is not clear that increasing the speed will lead to greater justice being done.

A further risk of highly specific actionable indicators is one of “teaching to the test”, or worse, “reform illusion”. The particular things that governments or aid agencies choose to measure might be areas amenable to quick action. But these actions may not be mirrored in other – rather important – areas not specifically covered by such “actionable” indicators, and thus such partial actions, while subject to “actionable” measurement, may not end up making a significant difference on outcomes.

Thus, there is a need to focus on ‘action-worthy’ indicators instead, ensuring not only that the indicator refers to actions that are likely to really matter, but also that the set is sufficiently comprehensive to avoid ‘tunnel vision’ focusing on easy actions (‘low hanging fruits’) – leaving pending many difficult reforms which are crucial for impact.

Further, since it is impossible to fully measure and monitor periodically all relevant ‘action-worthy’ indicators (let alone ‘actionable’), and also since the links between such action-oriented indicators (even if measured) and outcomes on the ground are far from certain, it is imperative to continue to also rely on outcome indicators – and in particular to continue to measure what is taking place de facto on the ground, according to the views of citizens, firms, and experts. Thus, there is also another important complementarity, namely between ‘action-worthy’ and outcome indicators, which also needs to be further emphasized and exploited, rather than moving away from measuring outcome results on the ground. This is particularly the case when “actionable” indicators are constructed by policymakers who then subsequently are reluctant to make available for public scrutiny. For example, it is striking that although the PEFA indicators have currently been compiled for around 30 countries, less than a third are publicly available, with adverse consequences for the transparency and credibility of the entire process.

New integrated governance tools, such as the in-depth in-country Governance Assessment being currently being implemented for Kenya, illustrate the benefits of exploiting complementarities between action-worthy and outcome indicators, as well as between individual and aggregate indicators, and between subjective and objective indicators. First, it is the preponderance of evidence emerging from many different types of indicators – rather than merely one source –, relying on the WGI (summarizing dozens of sources) as well as on individual international as well as local surveys, which leads to the finding of a sizeable initial improvement in

GOVERNANCE

BASEL INSTITUTE ON GOVERNANCE

Page 5

governance and in corruption control in Kenya shortly after the regime change takes place in late 2002, and to the subsequent deterioration – particularly over the past couple of years.

Second, the use of particular individual sources in a disaggregated way has shed light on the evolution of particular dimensions of governance, showing for instance reversals in the early progress in judiciary and procurement corruption. Third, it is by using specific indicators which are ‘objective’ (such as Doing Business) as well as ‘subjective’ (data from responses from surveys of firms), that one can assess in a more integrated fashion the particular challenges in specific areas – such as in the regulatory and licensing environment to enterprise, and also more fully take into account what matters the most from the perspective of Kenyan firms and investors themselves.⁸

Such an integrated empirical approach to assessing in depth governance within a country, drawing from the responses from thousands of citizens, experts and firms, complemented by objective de jure data, provides a potentially richer perspective for analysis than when governance was subject to lengthy prose by one or few desk writers without recourse to data, or solely relying on official or objective data. Concluding Thoughts In sum, in moving forward and improving upon all efforts on governance indicators, we would emphasize the following seven points:

- paying particular attention to the existence, measurement, and also to the implications of margins of error in all efforts to construct governance indicators
- full disclosure of margins of error and of sources and methodological details (including on data collection) in all indicators efforts, and in particular, emphasizing the key role of public access to the resulting indicators to ensure full transparency and credibility.

- utilizing more effectively and intelligently (with an appropriate abundance of caution) the myriad of indicators already in existence, while continuing to strive to improve upon them, as well as generating new indicators
- making progress in gathering data on a set of “action-worthy” indicators, which are not simply “actionable” or overly narrow in scope
- given the difficulties in linking particular actions with governance results on the ground, to continue to utilize and improve upon outcome indicators as a key part of the governance diagnostic process
- avoiding to the extent possible utilizing any one governance indicator alone for any major policy or aid decision-making – without applying checks and balances provided by other indicators
- finally, tying the various specific threads: the importance of exploiting the complementarity among different types of indicators – be they aggregate or individual, objective or subjective, action-worthy or outcome.

⁸ For details on the ongoing Kenya Governance Assessment, contact the authors.