Assessing the Progress of Anti-Corruption Efforts: "Actionable" Indicators of Reform

Michael Johnston
Charles A. Dana Professor of Political Science
Division Director for the Social Sciences
Colgate University
Hamilton, NY 13346

November, 2006

REVISED DRAFT:
All comments welcome

I. Introduction

Much effort and debate have gone into a variety of efforts to measure corruption—that, as part of a more general movement advocating the measurement and comparison of diverse aspects of governance (for an excellent account of the sources of that movement, and a listing of data sources on governance, see Arndt and Oman, 2006: Ch. 1, 2; on Latin America, see Transparency International, 2006; see also the United Nations Development Program Governance Indicators Project[1]). The resulting indices have been of indisputable value in a variety of ways—not least for helping direct public and official attention to places and regimes where governance is in dire need of improvement. But as we have moved, over the past several years, from laying out the seriousness and consequences of corruption to the longer-term challenges of reform, it has become clear that single-number corruption measures do not help reformers attack the problem in tightly focused ways nor assess their progress when they have done so (an excellent collection of articles reviewing these issues from a variety of viewpoints appears in Sampford, Shacklock, Connors, and Galtung, 2006). In this paper I will suggest that the challenge is not to measure corruption across whole societies, but rather to develop easily-gathered, transparent indicators of specific effects of corruption and incentives that sustain it. These sorts of indicators, I suggest, are not only far more focused upon specific agencies and levels of government, official processes, and segments of societies; they are better suited to tracking changes over time. Most important of all, perhaps, such indicators can and should be *actionable*—that is, should directly measure processes, incentives and risks we can change and for which accountability can be clearly specified.

[1] http://www.undp.org/oslocentre/cross.htm (Viewed 7 November 2006).

Why not use corruption indices?

While corruption indices have become increasingly sophisticated and broad-based over the past decade, from the standpoint of those charged with reform significant problems remain. Some of these issues will be discussed in more detail below, but a few are worth noting from the outset. One is definitions: consensus on a nominal definition of corruption has not been attained, and likely will never be reached. Another is that corruption is by nature a hidden phenomenon: generally, all who know of a corrupt act have an interest in concealing it, and there is usually no immediate "victim" with a reason to file a report. Most of the indices we have are based, to varying degrees and via differing sorts of calculations, upon perceptions. Perceptions are not the same thing as corruption itself—indeed, in some cases they can be misleading—and are vulnerable to "echo chamber" dynamics encouraging those who are asked to assess levels of corruption end up repeating "what everybody knows". Thus, it is hard to say what "a high level of perceived corruption" means. Many of the strategies commonly used to deter or reveal corruption—bookkeeping and official records, official powers of supervision and intervention, and indeed any reports that may be filed by citizens and whistleblowers—may, in the wrong hands, be used to conceal corruption rather than reveal it, to mislead investigators, or to use corruption allegations and accompanying punishments as weapons in political struggles. Further, some of the indices can do little to track trends over time, as they are affected by the number and identities of societies being compared, and because underlying perception data may be gathered at long or irregular intervals and are in some instances re-used for several years. Worst of all is *the single-number problem:* differing types of corruption occurring in different segments of societies and

governments are compressed into a one-number assessment of whole countries. Comparisons of countries along the one-dimensional indices that result in effect assume that corruption is the same thing wherever it occurs, varying only by degree, while statistical explanations for index scores apply the same substantive explanations everywhere.

In fact neither corruption nor reform is a single national-level process. Each is a series of complex, highly specific, and often interactive steps to alter the ways wealth and power are pursued, used, and exchanged. Good governance does not reduce to an absence, or near-absence, of corruption—even if perception-based indices could accurately indicate such a state of affairs, which they likely cannot. It is, instead, a matter of changing behavior and of institutionalizing forces opposed to corruption and supportive of accountability. Reforn requires, at a minimum, detailed knowledge of the incentives and constraints affecting the actions of officials and citizens. At another level it is essential that reformers be able to show citizens, top political leaders, and public employees that corruption controls are effective. No single number, even if based upon data gathered at appropriate levels, can serve all of those purposes.

II. Measuring Corruption: Are We There Yet?

Other than the question of definitions, few issues have so thoroughly stymied analysis and reform as that of measurement. Types and amounts of corruption vary among, and within, societies. Theory tells us that these contrasts reflect political and economic influences, history, and culture, and in turn affect societies and their development in

important and contrasting ways. But the difficulty of measuring corruption has long made it difficult to make such comparisons, to test hypotheses, and to build sound, comprehensive theories.

Problems of measurement

The problems start with an old dilemma—that of definitions. If we study corruption at a general level it might make sense to examine the core cases and not worry much about the boundaries of the concept. But when it comes to counting and measurement, and to keeping up with changes and mutations in corrupt practices themselves, the margins become critical—and there is much disagreement as to where those boundaries fall (see, for a summary of the definitions debate, Johnston, 2005a). Add to this the complex relationship between corruption and scandal (Moodie, 1980; Markovits and Silverstein, 1988) and it becomes clear that perceptions and corruption may be very different things. Public reports and controversies may tell us more about the *appearance* or openness of corruption—and thus, about conflicts among elite factions, political divisions in society, or journalistic practices—than about its actual extent.

Other difficulties are methodological. In principle we can measure anything (Babbie, 1995: 110). But that is more easily said than done, and it is a long way from essential concepts and nominal definitions to the events or artifacts included in operational measures. Many concepts are categorizations of, or inferences from, phenomena that are difficult to identify and observe. Consider "democracy" (Collier and Levitsky, 1997): we know it when we see it, but the concept remains essentially contested

(Gallie, 1965). Over time the concept has a way of "creeping" away from its starting point, necessitating a re-think of what it means (Collier and Levitsky, 1997)—or, in the case of corruption, to be re-defined *de facto* in terms of the most common practices, such as bribery, rather than assessed across the full range of variations and complexity. A more subtle danger is *reification* (Babbie, 1995: 116-118)—thinking about operational measures as though they were the concept itself. Given such difficulties it is not surprising that we often end up study things because they are easily counted.

Measurement becomes all the more difficult when that which concerns us is hidden. Where corruption is most serious the officials charged with control are themselves compromised; in such settings reporting corruption becomes an exercise in risk and futility. Violence or intimidation may be used to see off investigators and keep others quiet. Statistics on conventional crimes are notoriously inaccurate; how can we measure an activity that is usually clandestine?

Notable efforts to measure corruption

A variety of **corruption risk measures**, differing in breadth, methodology, and quality, are now available. Some of the longest-running efforts at measurement have been mounted by firms providing risk assessments to international business. These, some available on a proprietary basis only, have at various times included surveys by Political and Economic Risk Consultancy[2], the Institute for Management Development[3], Political

[2] http://www.asiarisk.com; see 2006 "Corruption in Asia Report" at http://www.asiarisk.com/subscribe/dataindx.html (Viewed 7 November 2006)
[3] http://www.imd.ch

Risk Services[4], *The Economist* Intelligence Unit[5], and Business International (now a part of *The Economist* group). Others are produced by advocacy groups such as the World Economic Forum[6] and Freedom House[7], survey organizations such as Gallup,[8] publications such as *The Wall Street Journal*, and groups of analysts, sometimes working in affiliation with international organizations. In the United States the Public Integrity Section of the US Department of Justice regularly publishes data on corruption convictions in federal courts. Economists have used measures of economic problems that, while not offered as corruption scales *per se*, tap into closely-related problems, such as data on "black-market premiums" or the quality of countries' institutions. A 1999 report by the United Nations Crime Prevention and Criminal Justice Division[9] compiling criminal justice data included statistics on bribery. The data encompass many countries and a long time span; on the negative side, reliance on official statistics raises questions of comparability across countries' definitions of corruption, court systems, and investigatory efforts.

The Opacity Index[10] (Hall and Yago, 2000; Kurtzman, Yago, and Phumiwasana, 2004), by contrast, gets at corruption by way of its correlates and consequences. It focuses upon "opacity"—"the degree to which countries lack clear, accurate, easily discernible and widely accepted practices governing the relationships among businesses, investor, and governments, and metaphorically the opposite of transparency—and its

[4] http://www.prsgroup.com; see data archives at http://www.countrydata.com (Viewed 7 November 2006)
[5] http://www.eiu.com
[6] http://www.weforum.org
[7] http://www.freedomhouse.org
[8] http://www.gallup-international.com
[9] http://www.uncjin.org/Special/GlobalReport.html (Viewed 7 November 2006) and the agency's *Global Report on Crime and Justice* (Oxford: Oxford University Press, 1999).
[10] http://www.opacityindex.com/opacity_index.pdf (Viewed 7 November 2006)

effects upon the open flow of information essential to efficient markets (Kurtzman, Yago, Phumiwaswana, 2004: 12). A statistical model incorporates indicators of corruption (Transparency International's Corruption Perceptions Index and the PRS Group's International Country Risk Guide—on the latter, see note 3), efficacy of the legal system, "deleterious economic policy", inadequate accounting and governance, and detrimental regulation—and estimates the net effect of such factors upon the interests rates paid by businesses operating in each of 48 countries. (The 2001 Opacity Index, by contrast, estimated interest premiums of discounts with respect to countries' sovereign debts). Not surprisingly, high estimated levels of corruption are associated both with unfavorable ratings on the Index (interest rates are higher countries with more severe corruption and related difficulties) and developmental difficulties in nations' macro economies. Neither the methods of computing the index nor, even more to the point, its constituent indicators are perfect, and coverage is limited to a relatively small number of countries. But the Opacity Index represents a notable innovation, and is worth careful consideration, because it is linked to real processes—flows of capital—that theory tells us should be quite sensitive to corruption and related issues, and that involve pragmatic evaluations and re-evaluations of governance on a minute-by-minute basis, as capital is borrowed and lent. Such evaluations are made by specialists and lenders who have a very large vested interest in getting things right, and who receive continuing feedback as to how accurate their assessments are proving to be.

Perceptions and realities

The most widely-employed indices are based on perceptions. Some rely upon sample surveys of the public at large; others depend upon expert assessments. People involved in international trade and investment are among those most often included in such surveys; given the lack of harder indicators, the fact that much corruption arises in the context of business deals, and the extent to which these people move about the global economy, this approach is a natural one. Moreover, perceptions of corruptness are significant in their own right, influencing foreign policy, aid, investment, and lending decisions. But the views of such businesspeople are of considerable interest they might well reflect a variety of influences besides direct knowledge of corruption. Further, surveys based on such business outlooks might well give mature market societies—where relationships between state and markets are relatively predictable and legally institutionalized—a kind of "pass", while ignoring a range of lower-level corrupt processes that might affect ordinary citizens. Few if any international business people show up at the U S Department of Commerce bearing satchels full of cash, while political contributions may be freely and legally given in the US—both facts that benefit the country's corruption index scores. By contrast, what does or does not happen on a dark country road in Louisiana—or New York—when a police officer stops a motorist is unlikely to figure into international corruption rankings.

Not all countries are included in all surveys; for some societies the survey database is rich and varied, while for others only a small number of soundings are available. Not surprisingly, sample sizes vary widely. So do the questions asked: some have respondents rate overall levels of corruption on a scale while others ask about

bribes, extortion, or other irregularities in specific governmental functions, or tap respondents' own experiences of corruption. Still others invite respondents to estimate the extent to which corruption or bribery harm a given country's economy, an issue that is not only complex in its own right but which also invites the conclusion that because country X is poor it must also be highly corrupt. Contrasting types of corruption, the stakes involved, the ultimate uses or destinations of corrupt benefits, effects upon specific segments of the population, and connections to organized crime or violence are just a few of the substantive variations that single-number comparisons omit. Results on such indices invite comparisons, but across a given year's results it is difficult to say what, if anything, is being compared, or what sorts of variations are either omitted or "flattened out" in the process of calculating results. Comparisons over time are even more suspect, for one year's score may be based on quite different evidence from another year's result (or, as is sometimes the case, a given survey might be kept in the calculations for two or three years running).

The best-known international corruption index is **Transparency International's Corruption Perceptions Index** (CPI),[11] launched in 1995 and updated annually. The CPI—a kind of "poll of polls" averaging out the results of a range of public opinion surveys, expert rankings, and other sources—has won worldwide attention, aided a variety of analytical studies, and—perhaps most important—has helped sustain public and official attention to corruption as a critical issue of development and governance. Coverage has expanded from 41 countries in the 1995 CPI to 163 in 2006, and

[11] Data, documentation, and press kit available at http://www.transparency.org/news_room/latest_news/press_releases/2006/en_2006_11_06_cpi_2006 (Viewed 6 November 2006).

methodology and documentation have been refined on a continuing basis (comprehensive discussion and justification of the CPI and its methods, by its creator, appears in Lambsdorff, 2006). Country rankings are based on averaged results from a minimum of three surveys, and from as many as ten in the cases of India and Indonesia. Over time the CPI has spawned some related indices and survey efforts by TI such as a "Bribe Payers' Index",[12] comparing the "supply-side" willingness of private parties in various societies to pay bribes elsewhere, and the *Global Corruption Barometer,*[13] offering more detailed evidence of public perceptions of corruption in key institutions of various nations (for a critique of the TI perceptions approach, employing data from the *Global Corruption Barometer* itself, see Abramo, 2005).

A more methodologically ambitious measurement effort are the **Worldwide Governance Indicators** (WGI)**,** part of the larger "Governance Matters" project compiled and published by the World Bank Institute.[14] A wide variety of indicators, including most of the survey data incorporated into the CPI but others as well, are used to estimate six attributes annually for over 200 countries and territories. In addition to control of corruption, those attributes include "voice and accountability", "political stability", "government effectiveness", "regulatory quality", and rule of law. The dataset also includes useful links to over thirty sources of constituent indicators, although some of the data figuring into the indices are proprietary and at least one source is confidential.

[12] http://www.transparency.org/policy_research/surveys_indices/bpi (Viewed 7 November 2006)
[13] http://www.transparency.org/policy_research/surveys_indices/gcb
[14] D. Kaufmann, A. Kraay, and M. Mastruzzi , 2006a. "Governance Matters V: Governance Indicators for 1996–2005", World Bank Institute paper; available, along with data and a variety of other sorts of documenatation, at http://web.worldbank.org/WBSITE/EXTERNAL/WBI/EXTWBIGOVANTCOR/0,,contentMDK:21045419~menuPK:1976990~pagePK:64168445~piPK:64168309~theSitePK:1740530,00.html (Viewed November 7, 2006).

The WGI data are calculated using a sophisticated "unobserved components" approach (Greene, 1990), and are notable for their extensive coverage and for breaking governance out into six substantively distinct components. Questions have been raised, however, as to what if anything the data actually measure, and about the ways errors are dealt with in the modeling and calculation processes (for one such critique, see Thomas, 2006). Arndt and Oman point out a variety of problems with the WGI indicators, identifying major difficulties in terms of error, comparisons over time, sample biases and a resultant "echo chamber" problem reminiscent of error issues identified by Thomas; and transparency of the data-gathering and computational processes that go into the indicators (Arndt and Oman, 2006: Ch 4; for a response to a variety of critics see Kaufmann, Kraay, and Mastruzzi, 2006b).

These ratings, and the scholarship and public debate they have spawned, have seemed to confirm much of what we had long suspected. Corruption rankings are worst for poor, undemocratic, and unstable countries. Multivariate analysis employing CPI data (and others) has produced solid evidence that corruption is both cause and effect of a variety of economic and political pathologies—connections sufficiently strong to place many developing societies into a kind of "corruption trap" (Johnston, 1998) in which economic and political problems are linked to corruption that contributes to further problems.[15]

[15] A very substantial bibliography emphasizing studies of the developmental implications of corruption, as well as one truly frightening photograph, can be found at http://people.colgate.edu/mjohnston/personal.htm

III. Problems of Measurement

Validity, reliability, and precision

All of the measurement strategies noted above involve the careful, well-documented use of systematic evidence; all have helped move the debate over corruption, its sources, and its consequences forward. But all are subject to difficulties that are more or less inherent in the process of measuring anything. Three such difficulties—or, three attributes by which we assess the value of any measurement—are **validity**, **reliability**, and **precision** (this discussion draws upon Babbie, 1995: 121-129). Briefly, *validity* refers to whether or not an indicator actually measures what we claim it does; ideally we want measures that are on-target and involve errors that are small and non-systematic. *Reliability* raises the question of whether or not repeated applications of a measurement tool or strategy, by a variety of users, return consistent results. A measurement tool that can be used consistently only by its author, for example, is less useful than one widely applicable by any knowledgeable individuals. *Precision* in a measurement is desirable, other things being equal, but can be more apparent than real; the best level of precision for expressing a measurement of anything might not necessarily be that involving the most decimal places. (Those less familiar with validity, reliability, and precision or seeking further discussion may wish to consult Appendix A of this paper).

Validity. Perception-based indicators are clear advances over the anecdotal evidence and hypothetical cases that dominated earlier phases of research, and over the diffuse and emotional claims often marking public discussions of corruption. Their results are broadly plausible: it is difficult to dispute the notion that Canada is less corrupt than

Poland, and that Poland is less corrupt than Kenya. In addition, the indicators relate statistically to others in ways that make theoretical sense—evidence for construct and predictive validity.

Problems arise, however, with the basic approach of using perceptions as our operational measure. Setting aside the difficulties inherent in measuring perceptions of *anything,* we must remember that perceptions are not the same thing as corruption itself (a good discussion of these and other validity issues appears in Miller, 2006). They may reflect the openness of corruption, rather than its actual extent. We can easily imagine one country in which corruption takes place openly, in myriad small-to-moderate transactions, and another where a few large, well-concealed deals at the top of the state structure—perhaps under the protection of the very officials and agencies nominally charged with bringing it to light—predominate. Visiting businesspeople who do gain access to such dealings might quickly acquire a stake in keeping their true perceptions to themselves. Corruption might distort politics, the economy, and development, and yet this country might score better on the CPI that its neighbor, where less serious corruption is practiced more openly.

Other subtleties complicate the rankings. What is being perceived as more or less serious? The many surveys contributing to both the CPI and WGI results yield results expressed in significantly different ways. Some produce perceptions of how corrupt a whole society is, while others deal with particular agencies or functions of the state. Various surveys ask about perceptions of the "problem", or of its "pervasiveness",

"level", or of "number of cases". Does extensive corruption reform to the number of cases, the sums changing hands, impact upon politics or the economy, or cases involving particularly important officials or programs (Rose-Ackerman, 1999: 4; Lambsdorff, 1999: 7-8)? Perceptions could reflect general impressions of whole societies, or expectations about the behavior of specific leaders or business figures. What appears to be an outbreak of corruption might actually be scandal stirred up by factional contention. Some judgments might reflect culture shock (particularly if one's basis for comparison is a low-corruption society), language limitations, or sheer dislike of a country or its regime. Surveys of citizens might tap into experiences with corruption—though such broad-based surveys are the exception rather than the rule—but again are vulnerable to differing social conceptions of corruption, "echo chamber" phenomena, variations in corruption across segments of society, and a range of likes and dislikes that might color perceptions of others' activities. The perceptions of outsiders—even if they rest upon a shared definition—might tell us little about the *significance* of corruption: what it *means* in its context. A seemingly minor case might be freighted with significance lost upon outsiders or ordinary citizens unfamiliar with elite conflicts. Some evaluators might be less than candid because of their own involvement in corrupt activities. Others who have not done well in business might exaggerate corruption to explain away their failures.

One-dimensional corruption indices, by their very nature, impose a single model upon corruption. But what is it? To the extent that the underlying model is coherent at all, it is that of bribery. Several of the components of the CPI specifically ask respondents to judge the extent of bribery, or of demands for bribes. Others implicitly emphasize

bribery by sampling business people instead of, say, poor farmers. (In that connection, three component measures ask recipients the extent to which corruption harms the business environment—confusing measurement with the question of consequences and inviting connections between corruption and broader economic problems.) Nepotism, official theft and fraud, *political* corruption such as patronage, so-called "petty corruption" such as police shakedowns of upon stall holders at local markets, and election fraud may not fit the bribery model (or the daily experiences of business people) so neatly, and may thus be underestimated.

Again, qualitative differences are collapsed into matters of degree. Bribery may be the main form of corruption in international business, and may be what springs to most minds when "corruption" is mentioned, but in some respects it is a special case. In a strict sense, bribery is a *quid-pro-quo* on *comparatively* free and equal (if illicit) terms. It differs from extortion, where officials force deals that may be anything but free and are rarely equal. Bribery seems most likely to dominate where corruption is moderate to moderately high, illicit deals are a matter of course, and participants are not frequently punished. Where the risk of punishment is high, or (by contrast) where powerful officials act with impunity, things may be different: in the former, bribe payers may have to add a "risk premium", while in the latter they are at the mercy of officials. In some corruption exchanges, such as patronage and nepotism, considerable time may elapse between receiving the *quid* and repaying the *quo,* and the two may be difficult to link or compare to each other. And other forms of corruption—electoral fraud, embezzlement, or using official resources to operate an under-the-table business—are not exchanges at all.

Reliability. Reliability is a relatively strong point of perception-based indices. Year-on-year correlations of CPI figures, for example, usually range from +.95 upwards, suggesting that one year's result for a given country is an excellent predictor of the next. Similar correlations between CPI and WGI corruption control scores are also strong—in the range of -.97 for some years, for example, with the negative sign reflecting the fact that effective corruption control receives higher scores on the WGI index while low perceived levels of corruption are given lower scores on the CPI (Johnston, 2005b). If these correlations were weaker or inconsistent we would have reason to doubt the CPI's reliability, but the consistency across time is striking.

But qualifications are in order, for year-on-year correlations within a given index could also be *too* strong: levels of corruption are likely to change, even if gradually, and to change in differing ways from one country to the next. A reliable scale should reflect these changes, too; thus, is a coefficient of +.94, for example, too strong? There is no real way of knowing. Moreover, some surveys included in one year's CPI have at times been carried over from previous years. While this broadens the base of societies that can be included, this method might also magnify the errors and biases in particular surveys while making the CPI's less responsive to real changes. Moreover, to the extent that actual levels of corruption are changing, all indices should ideally change in similar ways; but in practice *trends* in CPI and WGI data are correlated with each other at considerably weaker levels than are same-year results, and for a few countries seem to point to changes in opposite directions (Johnston, 2005b).

A related issue of reliability has to do with the degree of confidence we can have in the scoring. How sure can we be that a country perceived as having a high level of corruption actually has it in practice? While the architects of both the CPI and the WGI indices address reliability in this sense, the WGI project has been more forthright in dealing with the issue. CPI results are reported with a "confidence range" figure that is generally fairly tight. In fact it represents an estimate of a 90 per cent confidence interval—that is, a probability statement that a country's "real" score would be 90 per cent likely to fall between the upper and lower bounds of the "confidence range". But there is, by definition, a five per cent probability that a "real" score would be higher than the upper bound, and five percent likelihood that it would fall below the lower bound. And, since index scores are affected by the number and nature of the other countries included on the list—all must be ranked between zero and ten—it is difficult to say what the confidence interval really tells us. WGI indicators are not calculated on the basis of a finite scale; their mean is zero but the upper and lower limits of the distributions can and do vary from year to year. WGI's "unobserved component" technique (see, for a discussion of that approach, Greene, 1990: Ch. 17) treats country data as a linear function of governance—which remains unobserved, but is assumed to be normally distributed across countries—plus a "disturbance term". This approach, given certain assumptions, allows estimates of standard errors and confidence intervals—which turn out to be very large. Thus, a few countries can be placed at the good- and bad-governance extremes with a high level of confidence, but for most others the data do not clearly show that corruption control (or other attributes) is particularly strong or weak—much less allow

fine comparisons (Kaufmann, Kraay, and Zoido-Lobatón, 1999: 2, 15-19; see also Thomas, 2006).

Precision: The precision of the CPI and similar scales is difficult to evaluate. It is not obvious what units of measurement *any* corruption scale ought to use, or how we might expect observations to be distributed. Some surveys providing underlying data are anchored on absolute scales, while others are ordinal comparisons only (judgments that Country X is more corrupt than others, or that there are "a lot", "a few", or "no" cases of corruption among particular officials). Sample sizes, ranges, and distributions vary considerably, and thus sampling distributions and standard errors are likely to differ as well. Rendering these data comparable—and specifically, averaging ordinal-level comparisons with unknown, or possibly intercorrelated, error components, into a numerical ranking inevitably produces results shaped by the assumptions of the statistician. It is not clear that variations across all values—say, the difference between CPI scores of 5.0 and 6.0, versus 8.2 and 9.2—are consistent. The problem may be most difficult at the extremes—the high- and low-corruption cases that interest us most, and whose rankings draw most attention.

A different precision problem has to do with reporting results. CPI scores are reported on a zero-to-ten scale (with low scores referring to high levels of corruption, and *vice versa*) in tenths of points—and, for the 1995 through 1997 CPIs and all years' WGI scores, in *hundredths* of points. Is that sort of implied precision, even if it rests upon defensible methods of calculation, substantively meaningful? What would be an

appropriate level of precision? We have already noted the relatively large standard errors involved in WGI indicators, which at least are assumed to be normally distributed. Since the CPI does not have a true zero point, and if we are not certain that variations are consistent across all values, an argument can be made that it is essentially ordinal-level, and ought to be reported in broad bands (perhaps "low", "low-medium", "medium", and so forth) rather than in numerical points.

The single-number problem

Another issue has major implications not only for all three of our criteria—validity, reliability, and precision—but also for anti-corruption policy, is the "single-number problem" (for another critique of single-number assessments, see Arndt and Oman, Ch. 4 *et passim.*). Actual corruption varies in many ways: there are many forms (Johnston, 2005c) and contrasts within most societies. How much nepotism or patronage is equivalent to a certain level of bribery in road construction? Is that bribery comparable in significance to similar practices in arms contracting? No single national score can accurately reflect variations between Northern and Southern Italy, across Russia, or among Minnesota, Alabama, and New Jersey. Some countries have high-level corruption, others find it lower down the political or bureaucratic hierarchies, and still others see most abuses in electoral politics and patronage. It is easy to imagine a society in which one executive agency is deeply corrupt while another is much less so. In some countries the problem centers around international trade, while in others it is home-grown. And how can we track changes over time? Obviously any account of corruption, be it a case study or a data point, will be a simplification. But we might still ask how much

variation—quantitative and qualitative—*within* countries is obscured by assigning each a single number. Without more detailed knowledge, reform efforts and those who lead them will be flying blind, deprived of any but the most general or fragmentary evidence that their efforts are or are not bearing fruit.

What is likely to happen to single-number perception scores for a country that has begun to make meaningful progress against corruption? There are several possibilities: at the very least, progress will be uneven, and thus recognized more quickly by some observers than by others. In that event, the uncertainty (variance, or standard deviation in some versions) of scores might widen considerably while the scores themselves change in ways that would be difficult to interpret. More likely, a successful anti-corruption campaign would produce revelations of wrongdoing, convictions, and countless new allegations. This is all the more likely in a democratizing country with citizens, journalists, and opposition figures feeling more free to speak out, and contending factions using corruption allegations to settle old scores. In that setting, effective anti-corruption efforts might well cause perceptions to *worsen* markedly, at least in the short run. Finally, a campaign that begins to break up corrupt networks may well lead to a short-term surge of overt, smash-and-grab corruption as elites, uncertain about their hold on power, take as much as they can, as fast as they can take it (Scott, 1972; Knack and Keefer, 1995). Once again, ratings may worsen.

Corruption indices have helped maintain sustained attention for corruption issues, have moved old debates forward, and have framed new hypotheses for further work (but

for an argument that the TI CPI has outlived its usefulness see, Galtung, 2006). None has been proposed as the final word on measuring corruption; and, to discuss their weaknesses is ultimately to return to the inherent difficulties of measuring corruption. Still, the difficulties outlined above do matter. Existing indices likely help us least in the countries we care about most—those with the worst corruption problems. Even if country rankings make sense, causes, effects, and corrupt processes exist at several different levels of aggregation—contrasts that will not be revealed by single-number assessment tool.

IV. Measuring Reform, Not Corruption

How much guidance do corruption indices give reformers? Can those fighting corruption in a society look to CPI or WGI scores for evidence of progress, and for guidance in shaping their strategies? Can agencies practicing conditionality strategies, be they linked to outcomes or to progress over time, use index results to say with confidence which countries or agencies are succeeding, and which are not? In all likelihood they cannot. Such data exhibit impressive reliability in some respects, but reliability is not in itself evidence of validity. Moreover, results are not nearly detailed enough to guide reformers, and as noted above we do not know how well they track changes in levels of corruption. Perceptions may outrun, or lag behind, actual trends. Any comprehensive anti-corruption strategy will likely work better with some varieties of the problem than with others, and yet a single-number index will not be able to tell us much about those contrasts—and thus, much about where to attack first or which aspects of a strategy are working and which are not. These are much more than mere methodological niceties. Perceptions of

corruption do shape important decisions, but the danger is that they will lead to an "echo chamber" problem in which officials and investors repeat what they hear from each other, in effect, and in which anecdotes and perceptions acquire false authority through repetition. Analysts—and even more so, reformers—need less subjective indicators of problems and progress (a very useful discussion of options of that sort appears in Arndt and Oman, 2006: Ch. 6).

Judging reform by government performance

Rather than attempt to measure corruption itself, a better approach is to use *indicators of government performance* to track trends in the *incentives to corruption* and in its *effects.* Such indicators can and should be *actionable,* an awkward word for a critical idea: as much as possible our indicators must directly reflect, and point to, aspects of governance over which we have significant amounts of influence and control—problems and trends about which we can *do something*. It follows from that notion that our indicators must track changes in sensitive and accurate ways, not only giving reform leaders invaluable feedback on the effects of their efforts, but also—critically—enabling them to take credit for success or progress, or fixing accountability for failure.

Consider, as one example of such an indicator, bureaucratic delay. It is both an incentive to corruption and one of its effects. Assume that in City A getting a building permit involves 33 steps and takes seven weeks, while in City B the same process involves 4 steps and takes three days. We cannot measure corruption in those two agencies directly, but theory and anecdotal evidence suggest that the numerous steps and

long delays in City A are both effects of corruption (bureaucrats have found they can make money by contriving new requirements and delays) and incentives sustaining it (construction firms, knowing that time is money and facing numerous bureaucratic "toll gates" and long delays, will find it quicker and cheaper to pay up). In City B, by contrast, officials have fewer opportunities to extract payments and contractors have less reason to bribe. Quite likely there is less corruption in City B's agency; further, if City A reduces the number of steps and amount of time involved in its process we can infer (with caution) that corruption will be reduced there as well.

Information on the number of steps and time required for such routine government functions can be gathered relatively easily on a regular basis. The results are *verifiable*: an agency can compile them from its own records, while anticorruption agencies and citizen groups can file applications of their own, and can send individuals through various administrative processes, in order to gather their own data. Results can be published regularly as a kind of "governance barometer", and can be compared over time and among jurisdictions. The gathering of performance indicators can be targeted upon agencies, public functions, and specific communities where reforms are underway or contemplated, and can be used in pre-reform phases to identify priority targets.

This sort of monitoring is a good-governance activity in its own right. Developing the institutional capacity and will to gather and publish such indicators improves the management of agencies and the supervision of employees, and can be done without imposing extensive new administrative burdens (or creating new delays) in the

name of reform. Cutting the number of administrative steps and time involved in a permit or licensing process—to continue with the earlier example—can both indicate progress of reform *and reduce incentives and opportunities to corruption.* Publishing such indicators on a regular basis is a way to show the public, top executive and legislative leaders, aid donors and lenders, and potential bribe payers and recipients that reform is taking hold. Widely publicized and easily understood indicators offer reform leaders the opportunity to link their efforts to citizens' own problems and grievances, and to take credit (or blame) for results—a far more sustainable political basis for reform than simply calling for "political will." Finally, performance indicators are targeted hard data: they do not depend upon perceptions and do not reduce an entire country's diverse corruption problems to a single number.

Other Examples

The best indicators will be easily gathered, clearly linked to the effects of corruption and to its sustaining incentives, comparable from one place and time to another, and easily understood by citizens and non-specialists. A few examples might include:

- prices paid for fuel and basic clerical supplies
- prices paid for basic commodities involved in a public service (asphalt, concrete, vehicles, tools)
- prices paid to suppliers and charged from the public for basic services (school meals, telephone equipment)
- speed and accuracy with which vendors' invoices are paid
- time, number of steps, and range of variation involved in tax assessments and other revenue collections

- •time and charges involved in obtaining routine information, passports, or copies of documents
- •number of inspections performed per member of field staff in regulatory agencies
- •trends in the numbers of licenses, permits, and in subsidy or benefit payments granted by a given agency
- •citizen assessments of the quality and responsiveness of public services

There is a consistent logic underlying such indicators, and others: if one unit of government pays fifty per cent more for petrol than another; charges unusually high (or low) royalties for extracting natural resources; needs ten employees to accomplish tasks performed elsewhere by four; publishes budgets that are routinely far off the mark in terms of actual revenues, expenditures, and payrolls; has comparatively high rates of zoning and regulatory variances, or of tax exemptions and non-collection; conducts far more (or many fewer) regulatory inspections than other comparable units of government; or routinely receives poor marks for its public services from citizens or businesses, such facts point again to the effects of corruption and a range of incentives and gaps facilitating further abuses. Trends toward better performance—that is, toward benchmarks established by comparisons to other governments, or to open markets—cannot tell us directly that corruption is in decline, but do suggest that opportunities and incentives for it are being reduced, and that its effects (often closely related in practice to the incentives) are being reduced.

Comparisons among agencies and jurisdictions and over time are critical to turning such indicators into useful feedback for reform. It may be hard to say whether any given number of steps in a licensing process, for example, is high or low without

such comparisons. But once information from several agencies, gathered over time, is in hand reformers can watch for very low numbers (a deliberately sluggish process designed to extract bribes) or very high figures (frequent inspections as a form of bureaucratic or political harassment). Much the same logic applies to prices, expenditures, and other indicators.

This strategy is also of value to academic researchers. While it does not immediately broad-scale indicators of reform, composite indicators would not be difficult to produce. Some indicators (prices paid for basic supplies, for example, or time involved in obtaining common services) are more or less directly comparable across societies. Others could be standardized statistically. But the real rewards would be realized when we turn our study of corruption and reform to more detailed cases within societies and governments. There, we can and should be studying *different kinds* of corruption problems, and we can track those problems using different sorts of indicators. Any evidence that enhances our understanding of the incentives underlying corruption, or of its effects, will be of obvious value. To the extent that we are able to develop detailed and focused knowledge of cases *and* to compare those findings from place to place and over time, we may finally begin to develop and comparative, nuanced understanding of corruption that current cross-sectional approaches based on national-level corruption indices cannot provide.

Finally, indicators that are credible and easily understood will be valuable ways to build civil-society input into the reform process (for a discussion of "Citizen Report

Cards" in Bangalore, India, and of related ideas, see Thampi and Sekhar, 2006). Where citizens believe their views on the quality of services are wanted, are taken seriously, and over time appear to produce results, incentives to get behind reform are increased while perceived risks of doing so are minimized. Perhaps best of all, such indicators of improving services link to citizens' own self-interests as "consumers" of services, as taxpayers, and perhaps as property and business owners. Creating and encouraging such vested interests in reform helps break down critical free-rider aspects of reform. Where controlling corruption is an abstract ideal and a public good, few citizens will have compelling incentives to get involved in the process over the long run. But where reform begins to appear credible and to relate to self-interest, such participation will be much easier to encourage and to sustain.

V. Limits of this Strategy

Tracking trends in corruption and reform via performance indicators is an idea that will likely encounter considerable resistance in the very agencies, and among the administrators, that are of most concern to reformers. Top-level political backing and a judicious mix of pressure and support from aid and lending agencies will be required in order to get the strategy off the ground. A few targeted projects might be a good way to begin, particularly if they are followed up (where appropriate) by favorable publicity, enhanced job security, and perhaps higher salaries and status for successful administrators. Rank-and-file employees may resist as well, particularly if the monitoring is seen as unduly intrusive or as an early sign of impending job cuts. There too, rewards and positive incentives for successful and effective employees will be

critical, and will also reinforce broader good-government goals. A credible and skillful manager may be able to present this strategy to employees as a way to help them keep their jobs and to demonstrate their importance to the community and to elected officials.

Some kinds of data, even if they can be obtained, may be difficult to analyze. Such is particularly the case for procurement: while a liter of petrol is a liter of petrol, not all computers, office space, and consultant services are created equal. Comparisons will be difficult to make and interpret, and considerable creativity may be required in devising indicators. Given the sorts of resistance and pitfalls noted above, it may well be that many of these more challenging tasks of comparison should be deferred until the compilation, publication, and tracking of simpler and more "transparent" information has had some time to take root.

A further problem is obvious but worth noting: just because indicators are actionable is no guarantee that action will be taken. A certain amount of disruption and resistance to gathering even the most basic indicators, and to following through on them with effective controls, are inevitable. As monitoring efforts spread, *and in particular as they begin to put pressure on corrupt officials and their clients to significant degrees*, such resistance will only intensify. Being able to take credit for improving results, key aspect of the performance-indicators approach, may counterweight some pressures to some degree. But many leaders will find the upsets and uncertainties of effective monitoring and reforms threatening in personal terms—or, at least, less preferable than living a quiet life—and may do little to put the information to use. Outside pressure,

particularly (where appropriate) from aid and lending sources, may prove critical; and indeed the availability of good performance indicators will help make such pressures more precise and less arbitrary than they might otherwise have been.

The basic strategy proposed here will be less suitable and effective for some public functions and corruption problems than for others. As noted, it deals with routine and repetitive functions, and as such will not tap into large, one-time "grand corruption" transactions. There are also substantive and common-sense limits as to how streamlined a process ought to be, and to how many inspections a staff member should be expected to perform. Some agencies will deal with qualitative or non-repetitive decisions that are less easily compared with others—planning permission applications are an example. Still, it is not necessary to gather data on all public functions where corruption might arise. The point, instead, is to create a climate of positive incentives, public awareness, and continuous administrative scrutiny in which the quality of government performance takes high priority, and in which those who pursue such goals effectively can be rewarded for it, and to reduce opportunities and incentives to corruption in lasting ways.

Tracking trends in corruption and reform via performance indicators is just one piece of the broader reform puzzle. It may prove to be a controversial idea, in part because of its very effectiveness: gathering and publishing such indicators will likely encounter considerable resistance in the very agencies, and among the administrators, that are of most concern to reformers. Top-level political backing and a judicious mix of political pressure and support will be required in order to get the strategy off the ground.

A few targeted projects might be a good way to begin, particularly if they are followed up (where appropriate) by favorable publicity, enhanced job security, and higher status and (where possible) increased salaries for successful administrators. Rank-and-file employees may resist as well, particularly if the monitoring is seen as unduly intrusive or as an early sign of impending job cuts. There too, rewards and positive incentives for successful and effective employees will be critical, and will also reinforce broader good-government goals. A credible and skillful manager may be able to present this strategy to employees as a way to help them *keep* their jobs and to demonstrate their importance to the community and to elected officials.

The point, ultimately, is to create a climate of positive incentives, public awareness, and continuous administrative scrutiny in which the quality of government performance and service to citizens takes high priority. Monitoring performance indicators offers a more sensitive and useful approach than does using national-level perception-based data, and it offers useful feedback on the effects of reform at far less cost than strategies based on extensive household and business surveys. Over time this approach to tracking reform can help strengthen the links between citizen wellbeing and the ways in which public power and resources are used, enable those who govern effectively to be rewarded for doing so, and reduce opportunities and incentives to corruption in lasting ways.

Appendix A:
Validity, Reliability, and Precision

Validity raises the question of whether our data actually measure what we claim they do. Concepts themselves do not exist in the real world, or have "real definitions" (Babbie, 1995: 116; see also Hempel, 1952). They are, rather, constructs useful for categorizing objects or events, and for drawing out attributes we think they share. Thus our empirical measures can never be better than approximations, and the literature abounds with "measurements" that draw on something in addition to, or other than, that which they claim to measure—or that are grounded in nothing at all. As Babbie (1995: 127-128) explains, we can assess the validity of a measure in several ways. Does it possess *face validity*—that is, does it have anything to do with the concept we have nominally defined? An index that excludes extortion while counting street crimes might return higher values for places we think are more corrupt, but it does not measure what we mean by "corruption". Does it possess *criterion-related* or *predictive* validity, in the sense of predicting changes in other variables that theory tells us should be linked to our concept? For example, corruption measures should statistically "predict" the credit ratings lenders give to various governments. Or, a measure might be related to other variables in ways that are consistent with what we know about those factors, even if it does not "predict" them—an attribute called *construct validity*. We might, for example, expect extensive corruption where institutions are of poor quality (Knack and Keefer, 1995). A measure possessing *content validity* works well across diverse manifestations of a concept: corruption ratings ought to reflect the incidence of all the major varieties of corruption, not just one or a few. Finally, a concept might have *reference-group validity*—that is, be judged sound by people with extensive knowledge of whatever we wish to measure. This is of particular relevance to corruption measures, many of which draw upon the judgments of experts or international business people.

Reliability refers to the question of whether a particular measure returns consistent results. A corruption scale that rates Zimbabwe (say) as an 8 on a scale of ten one year, 2 the next, and 5 the year after that, is of little use: we have good theoretical reasons to expect that such wide variations are unlikely. No social-science measure will be completely reliable, but we can improve our results through careful construction of indices using good data, and by repeated testing.

Finally, **precision** refers to the fineness of the units in which our measure is expressed. In general, the more precision the better: we would have little use for a "yes/no" corruption variable. High-, medium-, and low-corruption categories would be better, and numerical rankings more precise yet. A related issue is **level of measurement**: some measures are *nominal*, grouping cases into categories among which there is no particular relationship (individuals' ethnicity, or the continent where a country is located, are examples). Others are *ordinal*, grouping cases into categories that can be ranked higher or lower in terms of some shared attribute. We might, for example, place countries into high, middle, and low GDP-per-capita groups; all in the "high" category would be more affluent than those in the "middle" group, but there would be considerable variation within groups and no assurance that the differences among groups are the same. *Interval-*

level measurements array cases along some common dimension demarcated by units of identical size, but without a point indicating the complete absence of the attribute. The Fahrenheit scale is an example: its zero point is arbitrary, so that while a one-degree difference is identical across all values, a reading of forty degrees is not twice as warm as a reading of twenty. We might survey residents of several countries asking whether officials are venal or public-spirited, and express the results on an interval-level scale (say, +5 to - 5). Such a measure could not, however, tell us a particular country has a total absence of public spirit or that it is twice as venal as some other. Finally, *ratio*-level data also array cases along a dimension marked off in units of identical size, but one possessing a true "zero point". Here, expressions of proportion are appropriate: a country with 50 million residents is twice as populous as its neighbor with 25 million.

Other things being equal, higher levels of precision and measurement are desirable. But there is such a thing as false precision: while it is more useful to know that a country's population density is 255 people per square mile than to say that it is moderate, it is neither useful nor statistically appropriate to express that measure as 255.348906346 people/mi^2. Paradoxically, one measurement can be more precise, but less accurate, than another: data telling us Country X's population density is 255 people/mi^2 may be less accurate than an ordinal ranking of "moderate" if the true figure is 75 people/mi^2. Level of measurement is an important statistical issue: it is tempting to treat ordinal data as interval-level, for example, but the results can be misleading.

References

Abramo, Claudio Weber. 2005. “How Far Go Perceptions?” Working Paper, Transparencia Brasil.

Arndt, Christiane, and Charles Oman. 2006. *Uses and Abuses of Governance Indicators.* Paris: Organization for Economic Cooperation and Development, Development Centre Series.

Babbie, Earl. 1995. *The Practice of Social Research.* Belmont, CA: Wadsworth, 7th ed.

Collier, David, and S. Levitsky. 1997. “Democracy with Adjectives: Conceptual Innovation in Comparative Research”, *World Politics* 49 (April), pp. 430-451.

Gallie, W. B. 1965. "Essentially Contested Concepts." *Proceedings of the Aristotelian Society* 56.

Galtung, Fredrik. 2006. “Measuring the Immeasurable: Boundaries and functions of (Macro) Corruption Indices.” Ch. 6 in Sampford *et. al., op. cit.*

Greene, William. 1990. *Econometric Analysis.* New York: Macmillan.

Hall, Thomas, and Glenn Yago. 2000. "Estimating the Cost of Opacity Using Sovereign Bond Spreads". Working Paper, Capital Studies Group, The Milken Insitute, Santa Monica, CA 90401.

Hempel, Carl G. 1952. "Fundamentals of Concept Formation in Empirical Science." *International Encyclopedia of Social Science II* (7).

Johnston, Michael. 1998. "What Can Be Done About Entrenched Corruption?", pp. 149-180 in Boris Pleskovic (ed.), *Annual World Bank Conference on Development Economics 1997.* Washington, DC: The World Bank.

Johnston, Michael. 2005a. “Keeping the Answers, Changing the Questions: Corruption Definitions Revisited.” In Ulrich von Alemann, ed. *Dimensionen politischer Korruption: Beiträge zum Stand der internntionalen Forschung.* Politische Vierteljahresschrift Sonderheft 35/2005. Wiesbaden: Verlag für Sozialwissenschaften, pp. 61-76.

Johnston, Michael. 2005b. *“Es possible medir la corrupción, ¿pero podemos medir la reforma?”* (“It is Possible to Measure Corruption, But Can We Measure Reform?”) *Revista Mexicana de Sociología* 2 (April-June), pp. 229-269.

Johnston, Michael. 2005c. *Syndromes of Corruption: Wealth, Power, and Democracy.* Cambridge and New York: Cambridge University Press./

Kaufmann, Daniel, Aart Kraay, and Pablo Zoido-Lobatón. 1999. "Aggregating Governance Indicators". Washington: The World Bank; August; http://www.worldbank.org/wbi/governance/working_papers.htm

Kaufmann, D., A. Kraay, and M. Mastruzzi. 2006a. "Governance Matters V: Governance Indicators for 1996–2005", World Bank Institute paper; available, along with data and a variety of other sorts of documenatation, at http://web.worldbank.org/WBSITE/EXTERNAL/WBI/EXTWBIGOVANTCOR/0,,contentMDK:21045419~menuPK:1976990~pagePK:64168445~piPK:64168309~theSitePK:1740530,00.html (Viewed November 7, 2006).

Kaufmann, D., A. Kraay, and M. Mastruzzi. 2006b. "The Worldwide Governance Indicators Project: Answering the Critics." World Bank Institute Paper (September). Available at http://web.worldbank.org/WBSITE/EXTERNAL/WBI/EXTWBIGOVANTCOR/0,,contentMDK:21045419~menuPK:1976990~pagePK:64168445~piPK:64168309~theSitePK:1740530,00.html (Viewed 7 November 2006).

Knack, Stephen, and Philip Keefer. 1995. "Institutions and Economic Performance: Cross-Country Tests Using Alternative Institutional Measures." *Economics and Politics* 7(3): 207–27.

Kurtzman, Joel, Glenn Yago, and Triphon Phumiwasana. 2004. "The Global Costs of Opacity: Measuring Business and Investment Risk Worldwide." *MIT Sloan Management Review* 46:1 (October), pp. 38-44. Available at http://www.opacityindex.com/opacity_index.pdf (Viewed 7 November 2006).

Lambsdorff, Johann Graf. 1999. "The Transparency International Corruption Perceptions Index 1999: Framework Document". Berlin: Transparency International (October); http://www.transparency.org/documents/cpi/cpi_framework.html

Lambsdorff, Johann Graf. 2006. "Measuring Corruption—The Validity and Precision of Subjective Indicators." Ch. 5 in Sampford *et. al., op. cit.*

Markovits, Andrei S., and Mark Silverstein. 1988. *The Politics of Scandal: Power and Process in Liberal Democracies.* New York: Holmes and Meier.

Miller, William L. 2006. "Perceptions, Experience and Lies: What Measures Corruption and What do Corruption Measures Measure?" Ch. 8 in Sampford *et. al., op. cit.*

Moodie, Graeme C. 1980. "On Political Scandals and Corruption." *Government and Opposition* 15(2): 208-222.

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

Sampford, Charles, Arthur Shacklock, Carmel Connors, and Fredrik Galtung, eds. 2006. *Measuring Corruption.* Aldershot: Ashgate.

Scott, J. C. 1972. *Comparative Political Corruption.* Englewood Cliffs, NJ: Prentice-Hall.

Thampi, Gopakumar K., and Sita Sehkar. 2006. "Citizen Report Cards." Ch. 12 in Sampford *et. al., op. cit.*

Thomas, M. A. 2006. "What Do the Worldwide Governance Indicators Measure?" Unpublished manuscript, Johns Hopkins University (October).

Transparency International. 2006. "Corruption and Governance Measurement Tools in Latin American Countries." Available at http://www.undp.org/oslocentre/docs06/Corruption%20and%20Governance%20Measurement%20Tools%20in%20Latin%20American%20Countries%20-%20June.pdf (Viewed 7 November 2006).