Measuring Corruption: perspectives, critiques, and limits

Paul M Heywood, University of Nottingham

Introduction: the current status of corruption measures

How do we measure something that is, by its very nature, largely hidden? This is the conundrum that faces all who have attempted to develop a means of measuring corruption. Given the seemingly intractable nature of this problem, the obvious question is why we should want to measure a phenomenon that is not only covert, but notoriously difficult even to define. There are, in fact, several reasons for doing so: first, it is important to assess the scale of the issue, in terms of its extent, location and trends, so that we know what we are dealing with. Second, we want to see whether there are any clear patterns in order, third, to help identify explanatory variables that will aid our understanding of why and where corruption develops. In short, measuring corruption will help us see better where we need to take action, as well as helping us decide both what that action should be and assessing whether it has worked. As we shall see, however, attempts at measuring corruption can lead to unintended consequences.

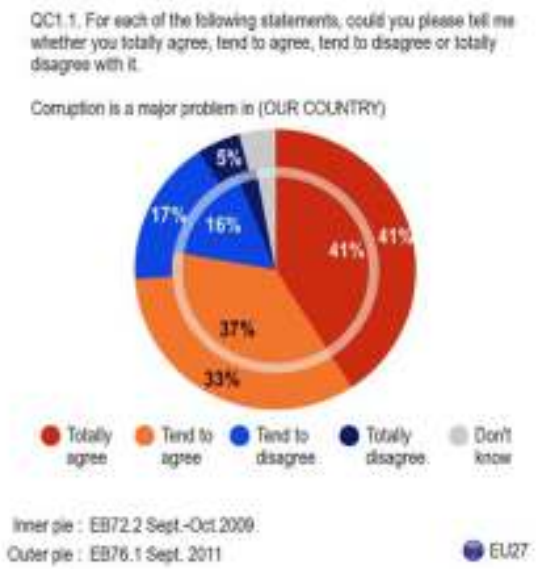
The dominant mode of measurement since the mid-1990s has been perception-based, via cross-national indices drawn from a range of surveys and 'expert assessments'. Indices such as the Corruption Perception Index (CPI), the Bribe Payers Index (BPI), the Global Corruption Barometer (all produced by Transparency International), the Business Environment and Enterprise Performance Surveys (BEEPS) or other aggregate indicators such as the Control of Corruption element in the World Bank Group's Worldwide Governance Indicators (WGI), have undoubtedly proved immensely important in raising awareness of the issue of corruption, as well as allowing for detailed cross country comparisons (TI 2009). However, it is now widely acknowledged that such measures are inherently prone to bias and serve as imperfect proxies for actual levels of corruption (Kurtz and Shrank 2007; Razafindrakoto and Roubaud 2006; Heywood and Rose 2014). Indeed, measuring corruption has been described as 'more of an art form than a precisely defined empirical process' (UNDP 2008: 8). Moreover the lack of an authoritatively agreed upon definition of what counts as corruption remains a serious obstacle to measurement, as in practice specific indicators inevitably (even if implicitly) reflect particular definitions which can be used to support different findings (Hawken and Munck 2009).

Perhaps the key stimulus to the dominant approach to measuring corruption has been Transparency International's Corruption Perceptions Index (CPI). First released in 1995 and published annually since then, the CPI has become established as the most widely cited indicator of levels of corruption across the world. The CPI 'captures information about the administrative and political aspects of corruption. Broadly speaking, the surveys and assessments used to compile the index include questions relating to bribery of public officials, kickbacks in public procurement, embezzlement of public funds, and questions that probe the strength and effectiveness of public sector anti-corruption efforts.' (TI 2010). The CPI is a composite index, calculated using data sources from a variety of other institutions (13 surveys and assessments released in 2011 and 2012 were used for the 2012 index). The CPI, though, has become increasingly controversial. Although widely credited with playing a crucial role in focusing attention on the issue of corruption, the index has none the less been subject to many criticisms both on account of its methodology and the use to which it has been put (see, for instance, Razafindrakoto and Roubaud 2006; Thomas 2007; Weber Abramo 2007; de Maria 2008; Andersson and Heywood, 2008; Hawken and Munck 2009; Heywood and Rose 2014). As is explicit in the title of the index, it measures *perceptions* rather than, for example, reported cases, prosecutions or proven incidences of corruption. This matters because perceptions can influence behaviour in significant ways: for instance, if we believe that all around us

people are engaging in corrupt behaviour, that may make us more likely to adopt such practices ourselves.

One of the recognised limits of aggregate perception data is that most factors that predict perceived corruption, such as level of economic development, state of democracy, press freedom and so forth, do not correlate well with available measures of actual corruption experiences (Triesman 2007). The potential scale of the disparity between perception and experiences of corruption is starkly shown in the regular Eurobarometer studies of the attitudes of Europeans to corruption (European Commission, 2007, 2009, 2012). For instance, the 2012 report, based on fieldwork conducted in September 2011, the study found that a strikingly high proportion of EU citizens (74 per cent average) saw corruption as a 'major problem' in their country, very similar to the levels found in the previous surveys (see Figure 1). In just five countries (Sweden, Finland, Luxembourg, The Netherlands and Denmark) did fewer than half the respondents agree. Those seen as most likely to be corrupt were politicians at national level, followed by politicians at regional level, then officials awarding public tenders and those issuing building permits – although personal experience of corruption remained very low, with just 8 per cent of respondents having been asked to pay any form of bribe for access to services over the preceding twelve months (European Commission 2012: 61).

Figure 1: Public views of corruption in EU member states



Source: Special Eurobarometer 374, *Corruption* (February 2012)

More generally, reflecting the same pitfalls in survey research beyond Europe, Triesman (2009: 212) cautions 'it could be that the widely used subjective indexes are capturing not observations of the frequency of corruption but inferences made by experts and survey respondents on the basis of conventional understandings of

corruption's causes'. A detailed study of the relationship between the CPI and TI's Global Corruption Barometer, which seeks to capture the lived experience of corruption through the eyes of ordinary citizens, has also shown convincingly that experience is a poor predictor of perceptions and that 'the "distance" between opinions and experiences vary haphazardly from country to country' (Weber Abramo, 2007: 6). Moreover, general perceptions cannot differentiate between various types of corruption, nor different sectors within countries. So, the question of whose perceptions, what their perceptions are of, and where those perceptions derive from is important.

Since the CPI is a composite index which draws upon a series of surveys mainly aimed at western business leaders and expert assessment, in practice the questions in many of the surveys relate specifically to business transactions (for instance, the need to pay bribes to secure contracts). Perceptions of corruption are likely to be seen primarily in terms of bribery, which cannot capture either the level of grand versus petty corruption, or indeed the impact of corruption (Kenny, 2006: 19; Knack, 2006:2; Olken, 2006:3). Moreover, the focus of questions is often on bribe-takers rather than bribe-givers: the implicit suggestion is that bribes are paid only when required by agents in the receiving country, rather than that they may be used proactively as a means to secure contracts.

A second widely remarked problem with the CPI relates to the question of how we can properly interpret what respondents to the various surveys understand by corruption. Each of the surveys operates with its own understanding of corruption (which may focus on different aspects, such as bribery of public officials, embezzlement and so forth), and seeks to assess the 'extent' of corruption (Lambsdorff, 2005: 4). However, although the surveys often ask a panel of experts to rank corruption on a scale of low to high (or some variation thereof), we cannot know whether the experts share a common assessment of what constitutes any particular location on such a scale: what seems a 'low/modest' level of corruption to one person, may look high to another (cf. Søreide, 2006:6; Knack, 2006: 18). In the absence of clear indicators, such rankings must be largely impressionistic. A third problem relates to the interval scales used in the CPI index, which since 2012 ranks on a scale of 100 (previously, it presented the scale as 1-10, to one decimal place). This suggests a high degree of accuracy can be achieved, and that a material difference can be identified between a country that scores, say, 70 and one that scores 67. That impression of accuracy is reinforced by the ranking being presented in a league table format – although, since the number of countries included in each CPI varies, the position in the table can be influenced simply by the how many countries are covered in any given year (see Knack, 2006: 20).

Although the CPI has been very important for research, there are other types of data – also based primarily on perceptions – that have been developed to some extent as a response to criticisms of the CPI. For example, Transparency International itself has published since 2003 the annual Global Corruption Barometer, based on a Gallup survey which seek to tap both into perceptions and lived experience of corruption, and the World Values Survey (approximately quinquennial since the early 1980s) includes questions on attitudes to corruption (e.g. Gatti et al., 2003). The World Bank's widely used Worldwide Governance Indicators (WGI) includes 'control of corruption' (identified as the exercise of public power for private gain) as one of six elements (Kaufmann et al., 2003, 2006) and is also a perception-based measure constructed through weighted averages and to some extent based on the same polls and surveys as the CPI (for examples, see Barbier et al., 2005; for a comprehensive critique of the WGI's construct validity, see Thomas, 2007).

Like the CPI, the WGI is a composite approach based upon a series of other indices: Control of Corruption (CC); Voice and Accountability (VA); Rule of Law (RL); Government Effectiveness (GE); Political Stability (PS); and Regulatory Quality (RQ). As

Apaza (2009: 140) has argued, the validity of applying the index rests on the ability of the WGI component indices to discriminate effectively among the six concepts, and to be different from other measures of government performance. Recently, however, using both measurement and causal models, Langbein and Knack (2010) have argued that upon closer scrutiny the six indicators are far from distinct (moreover most data users show no signs of familiarity with the underlying data). It is shown that while the indicators can provide a statistically reliable measure, 'what they reliably measure is not so clear (ibid: 365)' In fact, Thomas (2007) has argued that 'the constructs themselves are poorly defined and may be meaningless' and the UNDP (2008: 26) commented that 'by aggregating many component variables into a single score or category, users run the risk of losing the conceptual clarity that is so crucial.' If users are unable to understand or unpack the concept that is being measured, their ability to draw out informed policy implications is severely constrained.

The World Bank Institute's diagnostic surveys provide in-depth surveys of countries by using both experience- and perception-based questions, whilst the EBRD-World Bank Business survey asks more than 10,000 firm managers to estimate unofficial payments to public officials as a share of annual sales in firms 'like theirs' (although it is arguable that that these types of questions are not, as often claimed, indirectly experience based, since they ask how respondents perceive their surroundings rather than serving as an indirect way of reporting own experience – see Andvig, 2005). Finally, the International Crime Victim Survey asks respondents if government officials had solicited or expected bribes for service during the last year (see Svensson 2005). Since the mid-1990s, an increasing number of academic studies have begun to use these alternative measures of corruption either instead of or as a complement to the CPI. But many of these measures face the same problems of perception-based measures in general, and in the case of the widely used World Bank indicator 'control of corruption', the problems are very similar to those outlined above for the CPI (see Thomas, 2007).

Table 1: Summary of features of measures of corruption

Index/Survey Source	Definition of corruption measured	Information sources	Coverage	Interpretation
Corruption Perceptions Index (CPI)	Perceived corruption (composite) and some measures of corruption control	Statistical summary of expert assessments (e.g. expatriate business executives, senior business leaders, assessment by US, regional, and in-country experts)	Almost global depending on having sufficient sources. Annual (though not all data sources annual)	Cross-sectional ranking of perception of corruption focusing on business environment
Control of Corruption Index (CCI)	Perceived corruption (composite) and some business and public opinion survey evidence and corruption control assessment	Similar sources to CPI but with some survey evidence	Almost global depending on having sufficient sources. Biannual (though not all data sources annual or biannual)	Cross-sectional ranking of perception of corruption. Sources may be somewhat wider than business environment focus of CPI
Public Integrity Index (PII)	Overall institutional environment for controlling corruption	Expert assessment	25 countries	Absolute ranking (in principle allows assessment of change over time)
Bribe Payers Index (BPI)	Perceived willingness of companies from different countries to pay bribes, and sectors in which bribery most prevalent	Business experts	21 countries based on evidence from main emerging market economies. Last carried out 2002	Ranking of perceived willingness to pay bribes in different countries. Validity of perceptions and weighting uncertain
Global Corruption Barometer (GCB) and related surveys by Transparency International	Bribe payments by households and public perceptions of corruption prevalence	Public opinion surveys and partial household surveys	69 countries in 2005, though not nationally representative in many cases	Comparative prevalence and amounts of bribe payments though quality of survey data needs validation
World Bank Enterprise Survey (WBES)	Bribe payments by firms	Surveys of businesses	62 countries, various years	Quantitative comparisons of bribe prevalence and cost
Governance and Corruption Diagnostic Surveys (GCDS)	Bribe payments by households	Household surveys	16 countries	Quantitative comparisons of bribe prevalence and cost

Source: OPM 2007

Methodological issues

It follows from the above that for large aggregate indicators like the WGI, CCI, or CPI a gap can be identified between the concept and its measurement (Andersson and Heywood 2008; Langbein and Knack 2010). The cross pollination of assessment criteria, a lack of transparency, and data from different sources creates a tautological relationship between the dependent and independent variables, meaning that the indicators of the concept of corruption do not always relate systematically and reliably to how it has been defined (Langbein and Knack 2010: 351; Arndt and Oman 2006).

Hawken and Munck (2009) conducted an examination of the quantitative, cross-national literature on corruption that made use of the CPI and CCI between 1995 and 2009 – the first independent empirical assessment of the nearly full range of indicators used in corruption research (specifically 76 articles that appeared in prestigious economics journals) as well as the two most widely-used indices. The paper focused on two methodological choices. The first was the class of source used to generate data on indicators. Based upon the characteristics of the evaluator as the criterion of

classification, fives classes were identified: 1. Expert rating by commercial agency; 2. Expert rating by an NGO; 3. Expert rating by multilateral development bank; 4. Surveys of business executives; 5. Surveys of the mass public. It was shown that some evaluators are stricter than others, thereby generating a systematic margin of error (which reached as high as 14.7 per cent) both within and across countries and regions. Thus,

> 'As the analysis of indicators shows, a substantial amount of variation in reported levels of corruption is not attributable to variation in actual corruption or to random measurement error but, rather, is driven by the choice of evaluator and hence is an artefact of the method selected to measure corruption' (ibid: 12).

The second methodological choice was the aggregation procedures. The process of combining multiple (weighted) indicators was put forward as a way to reduce the measurement error of the individual indicators. Specifically, Kaufmann et al (2006; 2007) argued that by putting different individual indicators into common units, through a linear and additive aggregation rule, it is possible to measure corruption between countries whose data does not necessarily correspond in terms of time period or sector. However, this process 'hinges on the assumption that any error in the individual indicator is random as opposed to systematic and independent across sources' (ibid: 13). As Apaza (2009: 141) has pointed out, by collapsing different data sources – often selected only on the basis of convenience rather than theoretical justification – the aggregation model is unable to offer any nuance on the nature, category, or concept of corruption. As a result we cannot be sure of the underlying accuracy or what we are actually measuring. Therefore, even if consensus and high correlations exists between the CCI and CPI in the first place, this is by no means indicative of accuracy or validity:

> 'In a nutshell, data on corruption suffer from a fundamental problem, the fact that different data sets used in quantitative research are routinely associated with different findings, and that the relative validity of different measures of corruption and hence of the different findings is not readily apparent' (Hawken and Munck 2009: 2).

Nevertheless, the worldwide coverage offered by large datasets – a claim that can be made by Transparency International's CPI, Freedom House, and the World Bank Institute's CCI – has led to their widespread adoption by academics looking to test variables, the large-*n* cases offering a ready-made basis for analysis. As UNDP has noted, 'many of these same academics are critical of the methodologies used to generate these indices. Nevertheless, for academic users and researchers, the global coverage of data seems to trump data quality. After all, it is much easier and quicker to run a regression analysis using someone else's data, compared to the hard work of generating one's own' (UNDP 2008: 45).

Similarly, Urra (2007) also identified three problem-types that persist in the main aggregate measures of corruption (CPI, BEEPS, and WGI): 1. The perception problem; 2. The error problem; and 3. The utility problem. The perception problem is the large margin of error created when subjective indicators are used to produce complex statistical constructions that can easily create an illusion of quantitative sophistication. The error problem refers to both the internal margins of error already contained within the various sources of corruption data and errors relative to the concept itself – thus corruption research confronts not only sampling errors inherent to any social science research, but also the fact that any proxy for corruption must by definition be imperfect. The utility problem refers to the gap between measurement and solutions: the criticism here is that corruption assessments that are too broad are in turn difficult to convert into concrete anti-corruption initiatives. Azas and Faizur (2008: 11) argue that perception-based measures are actually antithetical as a means of combating corruption because

perceptions are strongly influenced by factors that have little to do with underlying realities. There is evidence that the CPI, for instance, acts as a 'lagging indicator', incorporating data that is two to three years old and is thus out-dated, especially in the face of burgeoning corruption scandals and/or prevention schemes and economic crises (Kenny 2009: 317). In addition, a government wanting to lower its corruption perception and that in doing so invites foreign experts and generates media attention about its efforts, does not necessarily combat corruption per se, but may just generate propaganda to change perceptions. In addition, such efforts can lead to a 'demonstration effect' whereby people emulate what are seen as practices that go unpunished, thus creating the impression that bribes must be paid, and it is alright to accept them in order to get things done (Cabelkova 2000).

Governance, democracy, development, and corruption

The data from TI's CPI suggest that GDP per capita correlates negatively with corruption, a statistical finding that has led to the widely accepted causal hypothesis that good governance leads to, or is a predictor of, economic development. Although this has assumed an almost scholarly consensus (Mauro 2004), it has undergone surprisingly little empirical scrutiny, an examination that once again calls into question the basic assumptions of measuring corruption (Kurtz and Shrank 2007). There is a potential problem of circularity when exploring the relationship between 'good governance' and corruption. A study by Kurtz and Shrank (2007: 539) of the WGI indicators has shown that those that seek to measure the probity and efficacy of bureaucracy are significantly coloured by recent economic performance and that perception-based measures are riddled with problems of adverse selection, and feature deeply entrenched biases for and against various public policy alternatives that are logically distinct from questions of public sector effectiveness.

> In fact, the contemporary paeans to public sector probity are so pervasive as to imply that the link between growth and governance is an article of faith or a starting point for analysis rather than a hypothesis subject to falsification (Kurtz and Shrank 2007: 538).

As the principal means of promoting democracy and development, as well as combating corruption, 'good governance' (GG) has become a catch-all epithet of the development community. In fact, concerns with governance and corruption emerged in the 1990s in response to the widespread failure of World Bank Structural Adjustments Programs (SAPs) and the loss of credibility of the so called 'Washington Consensus'. The criticisms, both economic and political, of the first generation neo-liberal reforms point out that governance and corruption 'provide convenient cover and an excuse for failure of policies not designed for development in the first place' (Azas and Faizur 2008: 13). This latter point perhaps pushes the case against the notion of corruption to a polemical extreme, however it also draws attention to the now inextricable relationship between development and efforts to measure and, therefore, control corruption. In fact, using the example of African corruption, de Maria (2008) has argued the TI's CPI can be used to subvert public administration to the agenda of Western economic interests. Termed 'neo-colonialism through measurement', it is argued that corruption cannot be comprehended outside the experience, nor can it submit to empirical investigation (ibid: 185). Whilst the CPI is perhaps 'oblivious to cultural variance', this type of critique is symptomatic of a post-structuralist 'critical turn' in the social sciences which tends to overstate the difference of the particular, thereby closing the analytical space for comparative and policy work (ibid: 188).

Unlike econometric indicators, which are commonly used to quantify and categorise developmental processes and outcomes, it is now widely agreed that corruption measurement requires much more elaborated constructions, subject to complex and, often subjective, inputs (Urra 2007). As shown above, a major criticism of corruption measures derives from biases in individual indicators, such as the perceptions of business leaders. For business people, good governance might mean low taxes and minimal regulation (e.g. free trade), whilst wider public demands might be for reasonable taxation and appropriate regulations (e.g. import inspections) (Apaza 2009: 142). Therefore where perception, policy and action meet, good governance can act as a euphemism for the free market, an idealised role for civil society that rarely exists in practice, and a clear separation of the bureaucracy from political influence – three factors that, when applied through various policies, can actually exacerbate underlying problems. Thus in situations where business people feel aggrieved by regulations and taxes, they may have a different evaluation of corruption compared to that of ordinary citizens.

Indeed, there is a paradox of development aid becoming increasingly conditional on the implementation of reforms that are impossible to achieve without that aid, hence generating the risk of a 'corruption trap' (Andersson and Heywood 2009). In light of this, it is possible to point towards an inherent politicisation of perception indices when (business) respondents with interests in a small non-interventionist state might report negatively upon states with stronger regulatory environments. This is not helped by the tendency for specific corruption studies to select their cases on the dependent variable, often not examining comparable cases in which corruption was less severe (Hopkin 2002, cited in Kurtz and Shrank 2007: 542).

The critical warning, therefore, is that 'links between governance and growth are thus more to likely to be artefacts of measurement than reflections of underlying causal dynamics' (Kurtz and Shrank 2007: 539). This has reportedly led to a diminished credibility of corruption perception measures in the eyes of many governments. A delegate at an international NGO reported that their personnel face problems working with governments because perception-based indicators fail to provide sufficient leverage to start a discussion on what needs to be tackled on the governance and anti-corruption agenda (UNDP 2008: 42).

It has also been suggested that, paradoxically, measuring the perception of corruption rather than corruption itself skirts the problem of measurement (Olken 2009: 2). Yet, this also raises the question of how those being surveyed form their perceptions in the first place and whether this correlates with objective conditions. Methodological interest has turned, therefore, towards the attempt to ascertain the accuracy of corruption perceptions, by correlating opinion-based surveys with objective studies. For instance, Svensonn (2003) conducted a study of bribe payments made by Ugandan firms using a unique quantitative data set combined with detailed financial information from the surveyed firms; Olken (2005) has constructed a 'missing expenditure' measure of a road building project in rural Indonesia by using engineers to estimate the prices and quantities of inputs in the road, and comparing this to official village expenditure and the perceptions of villagers themselves; Seligson (2006) collected data on corruption by using victimization surveys designed to gather information on specific government departments or officials by means of denunciation, where the questions in the survey invite the respondents to denounce corrupt acts and portray themselves as victims of corruption instead of active partners in corrupt transactions; and Ferraz and Finan (2008) have used external audits, released by the Brazilian government, to construct an objective measure of corruption based upon the number of violations associated with corruption. This allowed the authors to assess how the publication of incidents of theft or graft empowered voters to punish politicians at the polls.

Although there is no overall uniformity in the corruption measures deployed, what unites these studies is an attempt both to look beneath aggregate indicators such as CPI scores and GDP, and also to provide a more nuanced account of how reliable perceptions are as well as the social context in which they were formed in the first in place. As Kenny (2009: 329) has suggested, using 'output measures may better capture the development impact of corruption as survey data is likely to be more reliable than perceptions data, and so it is worth comparing the two to measure the accuracy of general perceptions scores as a proxy for corruption at the sector level.'

Notwithstanding this advantage, such an approach does not overcome the problems, outlined above, in terms of the internal limits of perception-based data. For instance, Olken (2006) finds that personal and community characteristics, such as ethnic heterogeneity, were significantly more correlated with corruption perceptions than were levels of objectively estimated corruption. Therefore the idea that it is possible to retrospectively test the reliability of perception indices by correlating that aggregate data with specific empirical findings is open to further question. Consequently, research developing innovative small-scale corruption measures does not offer greater theoretical sophistication, but trades conceptual work for rigorous empiricism. In part, this is perhaps because such work is informed by the notion, pervasive especially in neoclassical economics, that the main barrier to accurate measurement is not methodological, but information asymmetries. By improving the quality of data and grasping the limits and biases of the original perception indicators, more rational action can be induced within the development community. Yet, as it stands, the usefulness of such research 'is limited by the lack of a rigorous conceptual framework since it is not clear how to identify a corrupt act or how to generate an aggregate corruption measure' (Foster et al. 2009: 2). This suggests that research exploring both subjective and objective indicators is best suited to sub-national studies, a methodological caveat that precludes using the same strategy for national-level and wider comparative measures (Golden and Picci 2005). Indeed, given that most corruption takes place in local contexts, it is questionable why so many measures focus on the national level.

Development practitioners interviewed in a report conducted by the UNDP (2008: 41) report 'consistently that the most useful indicators are those that provide deep contextual information: Are there sufficient legal mechanisms to hold executive officials accountable for their actions? Are law enforcement officials paid appropriately? Are civil servants hired based on their qualifications and merits?' And in response to the paucity of international corruption measures that focus on poverty and gender issues, the UNDP has suggested that new indicators need to be developed at the national level. Questions like, 'how do poor women's experiences with corruption compare to poor men's,' need to be answered, but as it stands they are not (UNDP 2008: 29).

New directions in corruption measurement

As Hawken and Munck (2009: 21) have recognised, the task of measuring corruption, especially by developing cross-national data sets of broad scope, is laudable and welcome. Unfortunately, though, variations in reported levels of corruption are commonly a product of the prevailing methodologies as opposed to actual levels of corruption. The injunction they offer is to *know your data.* Thus despite the drawbacks, available data should not be jettisoned out of hand, but instead employed to generate a better index, through sensitive analysis of methodological choices on the basis of available data. In practice, this would require a narrower empirical scope, as both indicators and aggregation rules would first be based upon theory and rigorous tests; therefore,

> '...at this point it would be preferable to test theories about the cause and consequences of corruption with a smaller N than is provided by indexes such as the CPI and the CCI but with greater certainty that the data are more valid' (Hawken and Munck 2009: 24).

This shift would require a greater qualitative emphasis. One positive benefit of qualitative studies is the depth of insight they can offer. The main drawbacks, however, is that they can be bulky, hard to summarise, and difficult to compare across countries (UNDP 2008: 44). The United Nations Development Program Oslo Governance Centre (OGC) published *A User's Guide to Measuring Corruption* (UNDP 2008), commissioned from the NGO, Global Integrity, and produced in collaboration with the anti-corruption community, governance practitioners, researchers, policy makers, and donors. The guide explains the strengths and limitations of different measurement approaches, and provides practical guidance on how to use the indicators and data generated by corruption measurement tools to identify entry points for anti-corruption programming. Notwithstanding existing problems, it argues definitively that corruption can be measured. The solution offered is to 'employ multiple sources of quantitative data, qualitative narrative analysis and real-life case studies to "paint a picture" of corruption in a country, sub-national, or sector context' (UNDP 2008: 8).

Figure 1. Summary of Key Phrases

Assessments: Broad contextual analyses of the state and drivers of corruption that often rely on multiple indicators of corruption, including qualitative and quantitative corruption indicators.

Composite Indicators: A composite or aggregate indicator is one which combines different measures of a similar thing into a single measure. A well known example of this would be the Human Development Index which measures human development by combining indicators of life expectancy, educational attainment and income.

Corruption Indicators: Refer to discrete, often quantitative, measurements of a particular aspect of corruption (including the "level" of corruption).

Governance: Like corruption, the meaning of governance is manifold. For UNDP, it comprises the mechanisms and processes for citizens and groups to articulate their interests, mediate their differences and exercise their legal rights and obligations. It is the rules, institutions and practices that set limits and provide incentives for individuals, organizations and firms.

Objective Indicators: Indicators constructed from undisputed facts. Typical examples might include the existence of anti-corruption laws or the funding received by the anticorruption agency.

Perception-based Indicators: Indicators based on the opinions and perceptions of corruption in a given country among citizens and experts.

Experience-based Indicators: These indicators measure citizens' or firms' actual experiences with corruption, such as whether they have been offered or whether they have given a bribe.

Proxy Indicators: Buoyed by the belief that corruption is impossible to measure empirically, proxy indicators assess corruption indirectly by aggregating as many

"voices" and signals of corruption, or by measuring its opposite: anti-corruption, good governance and public accountability mechanisms.

Pro-Poor and Gender-Sensitive Indicators: A pro-poor indicator requires a focus on those living in poverty, and a gender sensitive indicator captures the different experiences and interests of women and men. Such indicators are useful to track the potentially different impacts that the mechanisms and processes of governance have on different social groups.

Input-based Corruption Indicators: Also called *de jure* indicators, these are indicators measuring the existence and quality of anti-corruption or governance institutions, rules, and procedures, i.e., the de jure rules "on the books."

Output-based Corruption Indicators: Also called *de facto* indicators, these are indicators that measure the impact of corruption on quality of life and public service delivery, i.e., the de facto deliverables of the governance system. These are difficult to precisely measure other than through proxy measures.

National Ownership: Refers to when local stakeholders, not outsiders, have driven and controlled the production of an assessment. Moreover, it is based on the premise of consulting with a broad range of national stakeholders, such as the government, civil society and the private sector.

Source: UNDP 2008

The guide suggests an 'informal taxonomy' that classifies corruption indicators into four categories: 1. The scale and scope of indicators; 2. What is actually being measured; 3. The methodology employed; and 4. The role that internal and/or external stakeholders play in generating the assessments. Given that the two terms are often conflated, as shown above, an attempt is made to distinguish between 'corruption' and 'governance.' The former is identified as just one, albeit important, aspect of governance making it necessary for users of indicator to understand where corruption stops and governance begins. In the UNDP report some of the professionals interviewed, echoing some of the points raised above, also insist that local indicators, developed in-country by domestic stakeholders rather than by international or external actors, should be the future of the corruption metrics field. These metrics are, by some standards, quite limited: they have little or no international coverage, are often purely qualitative, and may not be continued from year to year. But highly localized indicators that are customized to national or sub-national needs have the significant advantage of being designed from the beginning to yield actionable data (UNDP 2008: 43).

Along with the more established and widely used corruption indicators discussed above, a newer generation of measurement and assessment has emerged, like the Ibrahim Index of African Governance, the Global Integrity Report, and the Global Integrity Index. Joined by country specific indicators, this proliferation has raised criticism of a duplicative and distracting field that is in fact harmful to effective donor coordination and harmonisation of the reform agenda.

> Often the only thing that seemingly redundant measurement tools have in common is some combination of the words *governance*, *corruption, transparency, accountability*, or *democracy* (UNDP 2008: 12).

However, it is noteworthy that in May 2011, Global Integrity decided to remove from its website the Global Integrity Index which had ranked countries, citing as part of the reason that it was

> 'a conscious attempt to reinforce a key belief that we have come to embrace after many years of carrying out this kind of fieldwork: indices rarely change things. Publishing an index is terrific for the publishing organization in that it drives media coverage, headlines, and controversy. We are all for that. They are very effective public relations tools. But a single number for a country stacked up against other countries has not proven, in our experience, to be a particularly effective policy making or advocacy tool. Country rankings are too blunt and generalized to be "actionable" and inform real rebate and policy choices. Sure, they can put an issue on the table, but that's about it.' (Global Integrity 2011)

The emerging consensus in the field, therefore, is that disaggregated, qualitative, and internal/local assessments will more likely lead to actionable insights than composite, perception-based indicators. In an effort to move beyond inputs and outputs, researchers have also turned to political economy approaches that have been developed to understand what drives corruption in a country-specific context. This is premised on the notion that governance and corruption reforms are shaped by power relations embedded in social, political, cultural, institutional and historical contexts. However, 'A power analysis in development projects does little to help donors understand how to support and operationalise the findings'. Nevertheless, it is hoped that 'newer tools may help make such analyses more actionable in the future' (UNDP 2008: 25).

The UNDP report maintains that the panoply of corruption indicators can be complementary rather than inimical to each other. One set of indicators is not necessarily better or inferior to another – it depends what is being measured and toward what end (UNDP 2008: 36). The example used to support this claim is taken from Sierra Leone where statistical evidence (drug inventories from central government) pointed towards deep corruption as only 5 per cent of resources provided by central government could be accounted for at the local level. However, based upon perception surveys it was clear that systemic corruption in the healthcare system was not present in the minds of the public. In this case the perceptions of corruption, or outputs, did not match more objective measures of the sector's inputs. Such findings have raised further concern about the accuracy and usefulness of the methods used to measure corruption. The Pilot Project by conducted by the Hungarian Gallup Institute (HGI 1999) is often cited in the literature as evidence of the fundamental weaknesses of when different corruption measures are conflated. One of the central points made by the HGI is the methodological difference between the measurement of petty corruption (which would depend upon the perceptions of local people and provide a more accurate measure of corruption) and white-collar corruption in the higher spheres of state or business administration. Whilst the two are certainly not mutually exclusive, establishing a causal link is, in practice, highly problematic, even the though the latter can only exist within the established governance and social frameworks of the former.

In response, international agencies have pushed for 'actionable' indicators that measure specific features of corruption that are directly linked to policy decisions.

> 'To put it plainly, there is little value in a measurement if it does not tell us what needs to be fixed' (UNDP 2008: 8).

A possible alternative proposed by Johnston (2006) is to not to measure corruption across whole societies, but rather to focus upon the observable effects of corruption and the incentives that sustain them. Starting with specific agencies, different levels of government, and official processes would, it is argued, be better suited to tracking change over time. More specifically, some have proposed to measure corruption as the percentage of government officials willing to accept a bribe (Cule and Fulton 2005). However, a possible pitfall could be the instrumentalisation of action indicators. Trumpeting a particular policy area or sector can create a reform illusion, where direct

measurement of a particular area of corruption concern (e.g. civil service) is taken as a proxy for action, with concomitant effects on perception (cf. Heywood and Meyer-Sahling, 2013).

Also in response to the fact that the literature currently lacks a unifying framework by which different corruption measures can be placed together, Foster et al. (2009) have developed an 'axiomatic' research framework. Their approach seeks to deal with the plethora of available corruption measures with even greater mathematical sophistication through formal modelling, generating averages, of corruption indicators. This is said to provide a transparent methodology for classifying corruption measures, which, it is argued, can aid the researcher or policy-maker in choosing a measure and interpreting empirical findings largely because these measures 'generate additional insights and illuminate distinct dimensions of corruption that cannot be seen with the standard perception-based measures' (ibid: 20). The path down which this leads corruption research points towards further abstract formal modelling:

> 'Though our analyses are preliminary, we believe they are quite promising. Our methods of organizing data, constructing corruption measures, and specifying axioms, are readily implemented given appropriate data. They suggest additional survey questions that can improve the accuracy of results and their comparability over space and time. However, to assess whether a given comparison is statistically significant, or to test associated hypotheses concerning corruption, an additional set of statistical tools will need to be developed'.

Focusing on corruption in the Netherlands, Shacklock et al. (2006) argued that it is possible to assess the extent and prevalence of corruption by treating it as a specific subpart of a broad typology of integrity violations. The case is put forward to triangulate research, on municipal councils and police forces, by collecting and comparing information from different sources and methods, at different levels and in different sectors as well as at different points in time. Using the analogy of an iceberg, by descending below the surface the position of The Netherlands as an almost corruption-free country (in the top 10 of the CPI with a score of 8.7) is put under closer scrutiny. The next layer of the iceberg, *surveys of corrupt behaviour in the workplace,* provides an indication of nepotism, patronage, and cronyism and the bottom layer focuses on *self-reported* behaviour. This allows researchers to question the relationship between reputations and actual levels of corruption and the disjuncture between employees perceiving corruption and initiating internal investigations. Nevertheless, it also gives rise to a so-called 'integrity paradox' whereby greater vigilance and more numerous investigations may point towards greater corruption when in reality this may not be the case. This reflects the broader problem of never being able to bring all corruption to the surface, provoking the sanguine, but honest, conclusion that

> '[r]esearchers on corruption will have to live with the weight of the 'dark numbers'. We are exploring different parts of the iceberg in order to find out more about its characteristics as well as its extent. All presented methods have their problems as well as possibilities. All our research contributes to our knowledge about the complex and diverse nature of the corruption phenomenon' (ibid: 32).

Other country-focused assessments are provided by GRECO's (Group of States Against Corruption) country evaluations, the Working Party on the OECD Anti-Bribery Convention, and TI's National Integrity Systems (NIS) studies. All attempt to produce a country-specific analysis of problems relating to corruption and bribery, whilst operating with a broader template concerning the conditions under which corruption and bribery occur (Philp 2006). These reports have the advantage of uncovering issues which may be particular to states, thereby distinguishing elements of corruption that may be culturally distinctive or more widely shared with other societies (ibid.). Equally,

Transparency International has responded to the call for greater vigilance toward the behaviour of the West's business community by creating the Bribe Payers Index (BPI) to examine the 'supply side of corruption', focusing on the role of foreign firms from developed industrialized nations in offering bribes. This measure will go some way towards recognising the internal dynamics of corruption, rather than reducing it to a problem of developing nations.

Conclusion

This chapter has argued that the major corruption measures, such as the CPI, CCI, and WGI, which make use of cross-national perception indices to rank countries have been subjected to far-reaching criticisms. Concerns have been raised about both their methodological consistency and the political implications of the results they produce. It has been suggested that they all suffer from internal biases that may be more systematic than the creators, from TI and the World Bank, wish to admit. Nevertheless, the consensus is not to jettison such measures out of hand, but rather be more aware of their potential limitations when academics and development practitioners seek to build on this still valuable research data. The embrace of more objective hard measures, as opposed to subjective soft measures, has highlighted ways in which researchers can generate data on new indicators that are, in the first instance, disaggregated from proxies such as poverty, economic growth, and levels of democracy. This requires the setting aside of presuppositions generated by categories such as good governance, which, much like corruption itself, has no fully agreed definition. It is cautioned, however, that the difference between subjective and objective should not be overstated (Hawken and Munck 2009).

References

Apaza, C.R. 2009. 'Measuring Governance and Corruption through the Worldwide Governance Indicators: Critiques, Responses, and Ongoing Scholarly Discussion' *Political Science and Politics*, 42: 139-143

Andersson, S. and Heywood, P.M. 2009. 'The Politics of Perception: Use and Abuse of Transparency International's Approach to Measuring Corruption' Political Studies, 57, Dec: 746-767

Andvig, Jens Christopher, 2005. '"A house of straw, sticks or bricks"? Some notes on corruption empirics', *NUPI working paper* 678.

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

Barbier, Edward B., Richard Damania and Daniel Leonard, 2005. 'Corruption, Trade and Resource Conversion', *Journal of Environmental Economics and Management*, Vol. 50, No. 2, 229-446.

Cabelkova, I. 2000. 'Perceptions of Corruption in Ukraine: Are they correct?' CERGE-EI, Discussion paper

Camerer, M. 2007. 'Measuring Public Integrity', *Journal of Democracy* 17, 1: 152-165

Çule, M. and Fulton, M. 2005. 'Some Implications of the Unofficial Economy-Bureaucratic Corruption Relationship in Transition Countries.' *Economics Letters*, 89: 207-211.

de Maria, B. 2008. 'Neo-colonialism through measurement: a critique of the corruption perception index'*, critical perspectives on international business*, 4, 2/3: 184 – 202

European Commission. 2009. *Eurobarometer 72.2: Attitudes of Europeans towards corruption*, available at http://ec.europa.eu/public_opinion/archives/eb_special_en.htm (accessed 25 January 2010)

Ferraz, C., and Finan, F. 2008. 'Exposing Corrupt Politicians: The Effects of Brazil's Publicly Released Audits on Electoral Outcomes.' *Quarterly Journal of Economics*, Vol. 123, No. 2, pp. 703-45.

Foster, J. Horowitz, A. and Mendez, F. 2009. 'An Axiomatic Approach to the Measurement of Corruption: Theory and Applications.' Seminar Paper Midwest Economic Association Meetings, and the Walton College of Business.

Gatti, Roberta, Stefano Paternostro and Jamele Rigolini, 2003.'Individual Attitudes toward Corruption: Do Social Effects Matter?', World Bank Policy Research Working Paper No. 3122.

Global Integrity. 2011. 'Why We Killed The Global Integrity Index', at http://www.globalintegrity.org/node/792 (accessed 5 May 2011).

Golden, M. and Picci, L. 2005. 'Proposal for a New Measure of Corruption. Illustrated with Italian Data', *Economics and Politics*, 17, 1: 37-75.

Hawken, A. and Munck, G.L. 2009. 'Do You Know Your Data? Measurement Validity in Corruption Research' Working Paper, School of Public Policy, Pepperdine University

HGI, 1999. 'Basic Methodological Aspects of Corruption Measurement: Lessons Learned from the Literature and the Pilot Study', *Hungarian Gallup Institute*, December.

Johnston, M. 2006. 'Assessing the progress of Anti-Corruption efforts: Actionable Indicators of Reform', presentation at Workshop on Governance and Development, Dhaka, November 11-12, 2006.

Kaufmann, D. Kraay, A. and Mastruzzi, M. 2005. Governance Matters IV: Governance Indicators for 1996-2004. May, World Bank: Washington DC

Kaufmann, D. Kraay, A. and Mastruzzi, M. 2006b. 'Measuring Corruption: Myths and Realities,' Development Outreach, Sep, World Bank: Washington.

Kaufmann, D., Kraay, A. and Mastruzzi, M. 2006a. The worldwide governance indicators project: responding to the critics. World Bank Policy Research Working Paper 4149, World Bank, Washington, DC.

Kaufmann, D., Kraay, A. and Mastruzzi, M. 2007. 'Growth and Governance: A Reply', *The Journal of Politics* 69, 2: 555-62.

Kaufmann, D., Kraay, A. and Mastruzzi, M. 2008. Governance matters VII: aggregate and individual governance indicators 1996–2007. World Bank Policy Research Working Paper 4654, June, Washington, DC.

Kenny, Charles, 2006. 'Measuring and Reducing the Impact of Corruption in Infrastructure', *World Bank Working Paper* 4099.

Kenny, C. 2009. 'Measuring Corruption in Infrastructure: Evidence from Transition and Developing Countries' *Journal of Development Studies*, 45, 3: 314-332.

Knack, Stephen, 2006. 'Measuring Corruption in Eastern Europe and Central Asia: A Critique of the Cross-Country Indicators, *World Bank Working Paper* 3968.

Kurtz, M. and Shrank, A. 2007. 'Growth and Governance: Models, Measures, and Mechanisms,' *Journal of Politics* 69, 2: 538-54

Lambsdorff, Johann. 2003. 'How Corruption Affects Persistent Capital Flows,' *Economics of Governance* 4: 229-243.

Lambsdorff, Johann Graf 2005. 'Consequences and causes of Corruption – What do we Know from a Cross-Section of Countries' [online]. Available from: www.icgg.org/corruption.research_contributions.html [accessed 8 June 2007]

Langbein, L. and Knack, S. 2010. The Worldwide Governance Indicators: Six, One, or None? *Journal of Development Studies* 46, 2: 350-370

Manning, N. Kraan, D-J. and Malinska, J. 2006. 'How and why should government activity be measured' in Government at a glance? Draft, Project on Management in Government, Organization for Economic Cooperation and Development.

Mauro, Paolo. 2004. 'The Persistence of Corruption and Slow Economic Growth', *IMF* Staff Papers 51: 1.

Mauro, Paolo. 1995. 'Corruption and Growth', *Quarterly Journal of Economics* 110, 3: 681–712.

Olken, Benjamin A., 2006. 'Corruption Perceptions vs. Corruption Reality', *NBER Working Paper* 12428.

OPM, 2007. 'Measuring Corruption', *Oxford Policy Management* Briefing Notes 2007-01: www.opml.co.uk

Philp, M. 2006. 'Corruption Definition and Measurement', pp45-79 in Charles Sampford, Arthur Shacklock, Carmel Connors and Frederick Gatlung, *Measuring Corruption* Ashgate: London.

Razafindrakoto, M. and Roubaud, F. 2006. 'Are International Databases on Corruption Reliable?' A Comparison of Expert Opinion Surveys and Household Surveys in sub-Saharan Africa,' *Document de Travail,* 2006-7 Paris: Dial

Seligson, M. A. 2006. 'The Measurement and Impact of Corruption Victimization: Survey Evidence from Latin America.' World Development, 34, 2: 381-404.

Shacklock, A. Sampford, C. and Connors, C. 2006. 'Introduction', in Charles Sampford, Arthur Shacklock, Carmel Connors and Frederick Gatlung, *Measuring Corruption* Ashgate: London.

Søreide, Tina, 2006. 'Is it wrong to rank? A critical assessment of corruption indices', *Working paper* 2006:1, Chr. Michelsen Institute.

Svensonn, J. 2003. 'Who Must Pay Bribes and How Much? Evidence from a cross section of firms', *The Quarterly Journal of Economics*, Feb, 2003.

Svensson, Jakob, 2005. 'Eight Questions about Corruption', *Journal of Economic Perspectives*, Vol. 19, No. 3, 19-42.

Transparency International, 2009. Measuring International Trends in Corruption', *Transparency International*: U4 Anti-Corruption Resource Centre

Transparency International 2010. 'What is the CPI?', at http://www.transparency.org/policy_research/surveys_indices/cpi/2010/in_detail (accessed 20 April 2011).

Thomas, M.A. 2007. What do the worldwide governance indicators measure? Johns Hopkins University, School of Advanced International Studies, Washington DC.

Treisman, D. 2007. 'What Have We Learned About the Causes of Corruption from Ten Yeats of Cross-National Empirical Research?' Annu. Rev. Polit. Sci. 10: 211–44

UNDP, 2008. 'A Users' Guide to Measuring Corruption' *United Nations Development Program*, UNDP Oslo Governance Centre: Norway

Weber Abramo, C. 2007. 'How Much Do Perceptions of Corruption Really Tell Us?', *Economics Discussion Papers* 2007-19.