

Las colecciones de Documentos de Trabajo del CIDE representan un medio para difundir los avances de la labor de investigación, y para permitir que los autores reciban comentarios antes de su publicación definitiva. Se agradecerá que los comentarios se hagan llegar directamente al (los) autor(es). ❖ D.R. © 2002, Centro de Investigación y Docencia Económicas, A. C., carretera México - Toluca 3655 (km.16.5) ,Lomas de Santa Fe, 01210 México, D. F., tel. 727-9800, fax: 292-1304 y 570-4277. ❖ Producción a cargo del (los) autor(es), por lo que tanto el contenido como el estilo y la redacción son responsabilidad exclusiva suya.
21 de noviembre de 2002



NÚMERO 119

Arturo del Castillo.

**BUILDING CORRUPTION INDEXES.
WHAT DO THEY REALLY MEASURE?**

Abstract

The study of corruption increasingly has drawn on sophisticated statistical methods of causal inference. This is a welcome development, and the contributions of this quantitative literature are significant. However, with a few exceptions, quantitative researchers have paid sparse attention to the quality of the data on corruption that they analyze. The aim of this essay is to evaluate the methodological basis of the corruption indexes. Doing so, this essay attempts to demonstrate that these type of indexes are based on fairly ambiguous concepts and ad hoc methodologies. Therefore, the results from the growing number of quantitative works based on these indexes must be considered to be preliminary, although technically well done.

Resumen

Cada vez más, el estudio de la corrupción se basa en métodos estadísticos sofisticados de inferencia causal. Este es un desarrollo importante y las contribuciones de este tipo de literatura cuantitativa son significativas. Sin embargo, con pocas excepciones, los investigadores con una orientación cuantitativa han puesto atención en la calidad de los datos sobre corrupción que utilizan en sus indagaciones. El objetivo de este documento es evaluar las bases metodológicas de los índices de corrupción. Así, se intenta demostrar que este tipo de índices se basan en conceptos ambiguos y metodologías ad hoc. Por tanto, los resultado de los estudios de corte cuantitativo, que utilizan este tipo índices, deben ser considerados preliminares, aunque técnicamente bien hechos.

Introduction

Few issues have so thoroughly stymied the comparative study of corruption as that of measurement. Types and amounts of corruption vary among, and within, societies. Theory tells us that these contrasts reflect political, economical, historical, and cultural influences. But the difficulty of measuring corruption has made difficult to compare different cases, to test hypotheses, and to build comprehensive theories.

For many years, this problem was a concern mostly of academic analysts. But recently a variety of forces have put corruption back on the international policy agenda. These include, the globalization, growing competitiveness of the world economy, and the end of the Cold War, which reduced tolerance for corruption among ideological allies. Other influences include movements to ban international bribery by domestic legislation (the US Foreign Corrupt Practices Act), or by international agreements (the OECD Anti-Bribery Treaty, and the OAS Anti-Corruption Convention).¹

This revival of interest has spurred innovative attempts to measure corruption, often as a part of more general efforts of reform. However, these efforts often reflect the world-views of business and development interests. "Corruption" as an operational concept is becoming synonymous with bribery, and its impact judged increasingly in terms of economic development.² Few would dispute the importance of those concerns, but they have fashioned a new orthodoxy about corruption mirroring the broader "Washington consensus" over trade, aid, and development. With that has come a tendency for rich comparative concepts and findings to be overridden by a narrower vision treating corruption primarily as a problem of political and economical liberalization. The indexes and research that have resulted

¹ The increase in corruption since the advent of globalization is often cited as a reason for focusing more attention on corruption (Glynn, Kobrin and Naim, 1997). However, there is no evidence to support this claim of increased corruption over the past decade. Globalization may have increased our awareness of how much corruption impacts business. Increased competition in the global markets probably has had two effects that have drawn more attention to corruption. First, globalization has reduced the traditional spheres of influence of exporters of capital—most of all the United States. Success in a globally competitive market depends more on the firm's competence than on its national origin or the hegemonic powers of its government. Second, US managers have probably felt the disadvantage of the Foreign Corrupt Practices Act, which makes it illegal for US business to bribe foreign officials. Most likely, these changes brought about by globalization have raised demands from US business to level the playing field against other OECD firms on the one hand, and against domestic capital on the other. In a corrupt environment, domestic firms may have an advantage if they know the rules of the game better than foreign firms.

² Some could argue that the OECD governments' efforts are directed mostly at reducing bribery, not all forms of corruption. This may be motivated by the interests of OECD firms that operate abroad and are at a disadvantage against the domestic business in a corrupt system. Since operating within any system requires some familiarity with the "rules of the game", a corrupt system may discriminate against outsiders.

may draw upon detail knowledge of many people and groups, but ironically may also narrow our understanding of the problem.

The study of corruption increasingly has drawn on sophisticated statistical methods of causal inference. This is a welcome development, and the contributions of this quantitative literature are significant. However, with a few exceptions, quantitative researchers have paid sparse attention to the quality of the data on corruption that they analyze. The aim of this essay is to evaluate the methodological basis of the corruption indexes. Doing so, this essay attempts to demonstrate that these kinds of indexes are based on fairly ambiguous concepts and ad hoc methodologies. Therefore, the results from the growing number of quantitative works based on these indexes must be considered to be preliminary, although technically well done.

This essay considers the measurement problem in two main levels of discussion. One issue is the quality of the indices themselves. Another issue is their impact upon the policy and analytical debates. The purpose of the essay's critique is not to suggest that the new corruption scales are radically wrong. In fact, there is little reason to think they are. Nor is it to criticize the motives behind the various statistical indices. Rather, it is to emphasize the continuing need for a richly comparative and historical view of corruption, focusing upon many varieties of the problems and drawing upon diverse kinds of evidence and theory.

What Makes Corruption So Difficult to Measure?

In principle, social scientists ought to be able to measure anything (Babbie, 1995: 110). But this is more easily said than done. 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 may themselves be difficult to identify and observe. Consider "democracy": We know it when we see it, but the concept remains essentially contested (Collier and Levitsky 1997). Over the time the concept has a way of "creeping" away from its starting point, necessitating a rethink of what it means. Reaching consensus over definitions, let alone measurements, would be difficult. One result is that at times social scientists study things mostly because they are easily counted, not because they are really relevant for the academic or public debate (Babbie, 1995: 90-95). A more subtle danger is *reification*, it means, to think about operational measures as though they were the concept itself (Babbie, 1995: 116-118).

Measurement becomes more difficult when that which concerns us is hidden. We know corruption exists, but direct witnesses are few. More often, those with direct knowledge have an interest in keeping it secret. Where corruption is most serious the officials charged with control are themselves compromised, in such settings reporting corruption becomes risky. 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?

The issue is even more complex because the study of corruption lacks a single unit of analysis or theoretical model on which the empirical findings can be based. If we study corruption at a general level—particularly, if our concern is to study repeated syndromes—it may make sense to examine core cases of corruption and not worry much about cases at the margins. But, how to define where those boundaries fall? In addition, it is the methodological problem of how to operationalize the concept of corruption. Do we classify acts according to their degree of corruptness and count the actual incidence of corrupt acts? Do we use the number of public officials involved in corrupt transactions as well as the degree of corruptness of their actions? Or do we use the monetary value involved in corrupt transactions to measure the level of corruption?

Corruption Indexes: First-Generation Measures

A variety of corruption 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 assessment to international business. These include surveys by Political and Economic Risk Consultancy, the Institute for Management Development, Political Risk Services, The Economist Intelligence Unit, and Business International (now a part of The Economist group). Others are produced by advocacy groups such as the World Economic Forum and Freedom House. It is possible to mention also surveys produced by organizations like Gallup and The Wall Street Journal, which sometimes works in affiliation with international organizations.

Some of these measure efforts rely upon sample surveys of the public at large, or international business executives. Others depend on expert assessments. Not surprisingly, sample sizes vary widely. Some ask respondents to rate overall levels of corruption on a scale, others ask about bribes, extortion, or other irregularities in specific governmental functions.

Other sorts of data have also been used in the comparative study of corruption. In the United States, for example, the Public Integrity Section of the US Department of Justice regularly publishes data on corruption convictions in federal courts (Schlesinger and Meier 2000). Economists have used measures of economic problems that, while not offered as corruption scales *per se*, deal with closely-related problems, such as data on black-market transactions or the quality of countries' institutions (Knack and Keefer 1995). Another approach is the international compilation of criminal justice data by the United Nations Crime Prevention and Criminal Justice Division (United Nations 1999). These data encompass many countries and a long time span. However, these alternative sources of information have important problems for conducting accurate comparison due to the differences of the countries' definitions of corruption and court systems.

I. The Corruption Perception Index (CPI)

Most first-generation indexes measure *perceptions* of corruption, mainly from business people. Given the lack of harder indicators and the fact that much corruption arises in the context of international business, this approach is a natural one. Moreover, perceptions of corruptness are significant factor influencing foreign policy, aid, investment, and lending decisions. However, on the other hand, appearances can be deceiving.

This essay focuses primarily upon Transparency International's *Corruption Perception Index (CPI)*—the most widely used and, in many respects, the most ambitious effort to measure and compare perceived levels of corruption. The CPI—a kind of “poll of polls”—has won worldwide attention and aided a variety of analytical studies (Lambsdorff 1999b). Coverage has expanded from forty-one countries in 1995 to ninety-one in the 2001 version. Fourteen surveys are now used to calculate the CPI. At least three databases for individual countries are required for inclusion. CPI methodology has become increasingly sophisticated. Therefore, Transparency International (TI) publishes a comprehensive “framework document”, which explains the main technical aspects of the index (Lambsdorff 1999a). The CPI ranks countries on a scale from 10 to zero, according to the perceived level of corruption. A score of 10 represents a reputedly totally honest country, while a zero indicates that the country is perceived as completely corrupt.³

This index have seemed to confirm much of what we had long suspected. Corruption rankings are worst for poor, undemocratic, and unstable countries.

³ Before being added together, the indexes have to be standardized so that they all run from 10 (the least corrupt) to 0 (the most corrupt), whatever the original scale. To illustrate the principle, let one of the original scales run from 0 (no corruption) to 5 (most corrupt). Let us say that Mexico gets 4.0 on this scale. What should be Mexico's score on the CPI? First, we have to turn around the scale so that 5 becomes the least corrupt and 0 the most corrupt, and divide the absolute value of the CPI scale by the absolute value of the other index. In more complex cases, when for example the sub index of CPI does not contain the same countries, more complex procedures (as explained in Lambsdorff 1999b) have to be performed. In principle, the CPI index gives for each country each sub index an equal weight. Since some countries are covered by several indexes, each index will receive a lower weight for countries covered by many indexes. Moreover, in order to smooth the final CPI index several years of some of the sub indexes are included in the basis for estimating the 1999 CPI index. No countries where there are less than three observations, i.e. measurements from at least three sub indexes, are included. TI appears to be convinced that they have succeeded in constructing a successful index that is able to rank countries in a reliable way to the degree corruption is perceived to a problem. The basis for this claim is the high degree of inter-correlation between the 17 sub-indexes from which the CPI index is constructed (a correlation coefficient around 0.8 is common). Since some of the indexes with high inter-correlations are based on the information given by locals and others by expatriates or foreign experts, the bias coming from shared rumors or special experience of the expatriates is not likely to be serious, according to TI. Neither do, TI claims, any differences in the understanding of what is high or low corruption levels between locals because their understanding are highly correlated with the perception of indexes based upon the expatriates and foreign experts. Furthermore, these high inter-correlations are achieved despite the different ways the questions are phrased in the different surveys and polls. For details on the methodology, see: www.transparency.de

Multivariate analysis employing CPI data has produced solid evidence that corruption significantly slows and distorts economic development (Mauro 1997), and reduces foreign direct investment (Wei 1997). It is also linked to inflation (worse where inflation is high and variable –see Braun and Di Tella 2000), and weak political and administrative institutions (Knack and Keefer 1995). Corruption is marginally worse where political competition is weak (Braun and Di Tella 2000; Johnston 2000), and it is worse where ethnolinguistic divisions are severe (Mauro 1995; Easterly and Levine 1996).

Like any social-science measure, the CPI has strengths and weaknesses. Its value in sparking new research and public debate is beyond dispute. However, there is an important discussion about how to understand the CPI's ranking. TI has been careful to emphasize the CPI's limitations. Nevertheless it is important to underline some critical remarks. What it follows is a discussion about the CPI's validity, reliability, and precision.

a) Validity⁴

The CPI 2001 includes 91 countries. It is based on 14 different polls and surveys conducted by 7 independent organizations, not by TI itself.⁵ None of these surveys are dealing with corruption only, but they cover a number of issues of relevance for development and business confidence. TI, however, is using only the data on corruption. The majority of these indexes are based on fairly vague and general questions about the level or frequency of corruption perceived either by experts or business managers. About half are based upon expert opinions with in built checks to ensure cross-country consistency. The other half is mainly based on

⁴ *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). They are rather constructs useful for categorizing objects or events, and for drawing out attributes we think they share. Thus, empirical measures can never be better than approximations. As Babbie (1995: 127-128) explains, it is possible to 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 *predictive validity* –that is, does it have the sense of predicting changes in other variables that theory tell us should be linked to our concept? Corruption measures should statistically predict the variation of related concepts like the extent of transparency, or the credibility of the judicial system. Does it possess *construct validity* –that is, does it have a consistent relationship to other factors? For example, we might expect extensive corruption where institutions are inefficient (Knack and Keefer 1995) and there is a huge ethno-linguistic fragmentation (Easterly and Levine 1996).

⁵ The 7 organizations are: The World Economic Forum, The World Business Environment Survey of the World Bank, The Institute of Management Development, Pricewaterhouse Coopers, The Political and Economic Risk Consultancy, the Economic Intelligence Unit, and Freedom House. These organizations are partly non-profit development agencies, and partly consultancy companies specializing in strategic business information and market analysis.

questionnaires sent to middle and high-level management to either international or local firms.

The definition of the concept corruption also varies between the surveys. Thus, we may question whether the surveys cover the same phenomenon (see Lambdsorff 1999b). Furthermore, all the surveys ask for the extent of the phenomenon, although the meaning of “extent” is not obvious. Is it the frequency of corrupt transaction or the amount of bribes paid or money embezzled?⁶ Moreover, in general, the surveys do not distinguish between administrative and political corruption.

Each survey uses different sampling frames and varying methodologies. This may lead to inconsistencies between them. For example, the responses may depend on the respondents’ cultural background, and if they are residents or non-residents in the country in question. Furthermore, the responses may vary between income groups, among the experts and the general public. Lambdsorff (1999b), however, argues that the impacts of such factors on the CPI are insignificant for two reasons. Firstly, the correlation between the sources is high, which implies that the perceived “degree of corruption” is consistent among the different categories of respondents. According to Lambdsorff, this may cause the respondents to have the same idea of how to define “degree of corruption”. Secondly, even if the perceptions vary among the respondents, it still makes sense to aggregate the data and “obtain an assessment of the level of corruption seen by a broad and possibly heterogeneous sample of respondents”.

The CPI represents a clear advance in the empirical research of corruption. It makes possible to go beyond the anecdotal evidence and hypothetical cases that dominated earlier stages of the research on corruption. Its results are plausible. It is difficult to dispute the notion that Norway is less corrupt than Mexico, and that Mexico is less corrupt than Kenya. In addition, the CPI and similar scales relate statistically to others in ways that make theoretical sense.

However, problems arise with the basic approach of using *perceptions* as an operational measure. It is important to remember that perceptions are not the same thing as corruption itself. They may reflect the openness of corruption, rather than its actual extent. The two may differ considerably. Indeed, Rose-Ackerman (1996) has observed that as corruption problems become worse in a country, the major dealings tend to become fewer in number, and to take place closer to the top. It is possible to imagine one country in which corruption takes place openly, in small and moderate transactions, and another with less frequent, but large, well-concealed deals at the top of the state structure, perhaps under the protection of the officials and agencies nominally charged with bringing it to light. In the case where corrupt officials and their clients operate with impunity, informants and journalists might be silenced by

⁶ Note, that if the multiple equilibrium type of models is correct, there will exist large areas where the size of bribes will be positively correlated with corruption frequency (see Andving and Moene 1990). Hence, the distinction between size and frequency will not matter much for the construction and understanding of the index.

intimidation. In such a situation, business people (the main group targeted by the corruption surveys) might decide to keep their true perceptions of corruption to themselves. Corruption might distort politics, the economy, and development, and yet this country might score better in the CPI than its neighbor, where less serious corruption is practiced more openly.

Other problems complicate the ranking. What is being perceived as more or less serious cases of corruption? How much do judgements reflect levels of corruption? How much are they reactions to trends? Does extensive corruption refer to the number of cases or cases involving particularly important officials or programs? Perceptions could reflect general impressions or ethical expectations rather than knowledge of corruption as such. What appears to be corruption might actually be scandal stirred up by feuding factions (Rose-Ackerman, 1999). Some judgements might reflect culture shock, language limitations, or sheer dislike of a country or its regime.⁷

Another validity problem is the use of numerical rankings. Numerical rankings effectively treat corruption as a single generic process or problem, inviting statistical analyses that impose a common model upon widely varying societies and cases. However, as Luis Moreno Ocampo, a former Argentine prosecutor, has pointed out, “corruption in Sweden and Nicaragua are just not the same. To use the same word for completely different situations can only generate confusion” (Ocampo, 1993).

If corruption indices tend to impose a single model or type of corruption, what is it? To a significant degree it is that of bribery. Several of the components of the CPI specifically ask respondents to judge the extent of bribery. Others implicitly emphasize bribery by sampling business people instead of poor farmers. Nepotism, official theft, political corruption such as patronage and clientelism may not fit in the bribery model. Perhaps bribery is the main form of corruption in international business, and may be it is what most people have in mind when the word “corruption” is mentioned. However, in many respects bribery is just a special case of corruption. Bribery is the payment of a fixed sum, a certain percentage of a contract, or any other favor in money paid to the state official in charge of making contracts on behalf of the state or otherwise distribute benefits to companies or individuals, businessmen and clients (Del Castillo, 2001). Respondents to the CPI’s surveys may be well aware of the distinction between bribery and other kind of corruption, but their knowledge can not be showed in any single rating.

⁷ On the problems of the language of corruption see: Genaux (2000) and Moroff and Blechinger (2000).

b) Reliability⁸

Reliability is the strongest point of the CPI. The combination of several sources of information reduce the possibility of misinterpreting the responses from individual countries (Lancaster and Montinola, 1997). Furthermore, the surveys cover different countries, and each country in the CPI is not necessarily covered by all the 7 sources. To be included in the index, a country must at least be covered by three surveys from three different institutions. The surveys data must also refer to current conditions and not be more than three years old. Doing so, the data reflect the views of thousands of individuals who encounter corruption in differing ways in a range of countries, and are gathered in a variety of ways.

Given the links between corruption and basic political, economic, and institutional processes, a reliable index should return broadly consistent values from one year to the next. This is the case of the CPI. Table 1 presents the correlations among the CPI scales published for the period 1995 to 1999.

Table 1
Correlations Coefficients Between Different CPI Scales

	1995	1996	1997	1998
1996	0.9770 (41) P=0.000			
1997	0.9354 (42) P=0.000	0.9689 (47) P=0.000		
1998	0.9450 (42) P=0.000	0.9663 (53) P=0.000	0.9880 (52) P=0.000	
1999	0.9386 (42) P=0.000	0.9594 (53) P=0.000	0.9820 (52) P=0.000	0.9933 (85) P=0.000

Where: Coefficient
(Cases)
1-tailed significant

⁸ Reliability refers to the question of whether a particular measure returns consistent results. A corruption scale that rates Mexico, for example, as an 8 on a scale of ten one year, 2 the next, and 5 the next year after that, is a very poor reliable scale. Theory suggest that such wide variation are unlikely. No social-science measure is completely reliable, but it is possible to improve our results through careful construction of indexes using good data, and by repeat testing.

If these correlations were weak or inconsistent we would have reason to doubt the CPI's reliability, but the consistency across time is striking. However, coefficients 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. What does a coefficient of almost 0.94 between 1995 and 1999 scores really mean? There is not sure way of knowing it. Nine of the seventeen component measures in the 1999 CPI are actually three surveys taken in the same, or very similar, ways in three different years (1997, 1998, and 1999). This method although insulate the scores from short-term fluctuations caused by sensational scandals, it might also magnify the errors and biases in particular surveys. Thus, this method undermines the CPI's responsiveness to real changes. Comparability is also a problematic issue. Scores for countries with thirteen or fourteen surveys include most or all of the repeated measures –meaning that their scores reflect perceptions over several years— while those based on just a handful of surveys will not.

c) Precision⁹

The precision of the CPI and similar scales is difficult to evaluate. The CPI assumes that corruption is a one-dimensional phenomenon varying along a single continuum. Yet, corruption is not one-dimensional. Corruption has many facets, including embezzlement, bribery, and extortion. The CPI does not distinguish between these types of behaviour. Moreover, there are wide variations in the way corruption is organized, how the incomes from corruption are spent, and so on. These variations are likely to produce different economic outcomes. Neither does the CPI discern between grand and petty corruption. What the index does show is how systemic corruption is perceived by the chosen informants.

Since the CPI is based on perceptions, it does not necessarily reveal the true extent of corruption in a country. The CPI may also be biased against poorer regions of the world. For instance, we should not assume that a similar score means the same thing in countries at different levels of development. However, people's perceptions about corruption may be important for what actually happens. The mere belief that government officials are crooked may, for instance, affect business confidence and, in turn, investment behaviour (Goldsmith, 1999: 875)

⁹ *Precision* refers to the fineness of the units in which a measure is expressed. Indexes could be organized according to the "high", "medium", and "low" corruption categories, but also according to numerical rankings. A related issue to the concept of precision is the *level of measurement*. Some measures are *nominal*, it means, grouping cases into categories among which there is not particular relationship. For example, the continent where a country is located. Others are *ordinal*, it means, grouping cases into categories that can be ranked higher or lower according to some shared attribute. For example, place countries into high, middle, and low GDP per capital groups. In this case, all in the "high" category would be more affluent than all in the "middle" group, but there would be considerable variation within groups and no consistent variation among groups.

It is not obvious what units of measurement any corruption scale ought to use. While the many measures considered into the CPI contribute to its reliability, they 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. Indeed, what we mean by “more” or “less” corruption is less obvious than it may seem. This is a problem reflected in various surveys. Some ask about perceptions of the “problem”, while others ask about the “level” of corruption or even for “number of cases”. CPI architects argue that these are comparable assessments of the “degree” of corruption (Lambsdorff, 1999a: 7). However, it is possible to question this, particularly in differing linguistic settings.

Another problem are the rating scales used by the surveys. Some ratings are anchored on absolute scales, while others are ordinal comparisons only (judgements that country X is more corrupt than others, or that there are “a lot”, “a few”, or “no” cases of corruption among particular officials). The Freedom House ranking was not even expressed numerically in its original form. Sample sizes, ranges, and distributions vary considerably, and thus sampling distributions and standard errors are likely to differ as well. Rendering these data comparable inevitably produces results shaped by the assumptions of the statistician as well as by actual perceptions or events. One specific result of these difficulties is that while we often treat CPI data as ratio or interval-level, variations across all values—for instance, the difference between 5.0 and 6.0 versus 8.4 and 9.4—may not be consistent.

Closely related to this are differing lists of countries to which various component measures apply. Ideally we would have the same large number of corruption measures for every country, but we do not. The architects of the CPI have required a minimum of three corruption surveys before a country can be included—an approach that minimizes the error that might result from relying on just one or two ratings, but the missing data are not randomly distributed. Countries with poor institutions and governance also tend to have the fewest scales available. Therefore, those with the worst corruption might well have the least data, while others slightly better off, where at least some surveys have been conducted, may be wrongly viewed as the world’s most corrupt societies (Kaufmann, *et al.* 1999a: 22-23). TI regularly warns against interpreting CPI results in that way. However, variations in amounts and quality of data among countries raise validity and reliability problems.

A different precision problem concerns the reporting of results. CPI scores are reported on a 0-to-10 scale (with low scores referring to high levels of corruption, and vice versa). However, since the CPI does not have a true zero point and we are not certain that variations are consistent across all values, this scale is very imprecise. What would be an appropriate level of precision? An alternative is to report the score in broad bands (perhaps “low”, “low-medium”, “medium”, and so forth) rather than in numerical points. This alternative must include an accurate definition of each label of measurement.

Perhaps the most serious drawback of the CPI, and similar indices is what might be called the “single-number problem”. It is a precision issue, but one with

validity and reliability implications as well. Actual corruption varies in many ways. There are many forms and contrasts within most societies. When respondents judge the amount of corruption in a society, are they responding to the same things, or do their judgements reflect *qualitative* variations from one case to another? How much nepotism or patronage is equivalent to a certain level of bribery in road construction? In fact, no single national score can accurately reflect contrasts in the types of corruption found in different regions of the same country. For instance, is it the same kind and level of corruption in the Southern Italy than in the Northern Italy? Some countries have low-level of corruption in the bureaucratic hierarchies, but still have abuses in electoral politics and patronage. In other countries the problem centers around international trade, while in others it is home-grown. Obviously any account of corruption will be a simplification, and the CPI's architects have no control over the interpretations that result. However, we might still ask how much understanding is lost by collapsing complex variations—qualitative as well as quantitative—into single-number ratings.

II. *Why Do These Problems Matter?*

First-generation corruption measures have helped move the debate on corruption forward. At the same time, they have framed new hypotheses for further work. Still, the difficulties outlined above do really matter. They are important issues because the general public, journalists and other will often misinterpret the meaning of an index ranking, like the CPI. It is believed that the country scoring lowest on the CPI is the most corrupt in the world. This is a premature presumption. Firstly, the surveys used to construct the CPI only cover countries where the required information is available. Thus, a number of countries are not included, for instance Burundi, Cambodia, Laos, Nepal and Turkmenistan.

Secondly, even with a clear understanding of this shortfall, as well as knowledge of the methods used to estimating the indexes, it is still difficult exactly to say what a specific position in the indexes implies. One problem is that we do not have a standardized way of estimating the *level* of corruption. Should this level, for instance, be defined as the number of transactions affected by corruption or the total number of bribes during a year? Alternatively, should it be understood as the total sum of costs and benefits of corruption to the society, or the extent to which politics are affected by corruption?

Even if country ranking make sense, causes and effects of corruption exist at several different levels of aggregation. Thus, carrying out careful and nuance accounts of corrupt processes remains a central task for comparative analysis. Without such foundations, the significance of any ranking is open to debate. For instance, in 1999, Transparency International for Latin America and the Caribbean (TILAC, 1999) presented a study which argue another form to look at the CPI. This study emphasized the range of variation in ranking across the Americas and compared scores for this region to those of other parts of the world. The results

made Latin America corruption appears not quite as exceptional than we might have assume. This study obliges us to ask about the actual meaning of this kind of index. What does the term “extent of corruption” actually tell us? And how does the uncertainty around this issue affect the meaning of a ranking on a corruption index? It is tempting to think that a ranking of 6 on the CPI implies that the country in question is twice as corrupt as a country with the ranking of 3. However, this is not the case. Even without going into the actual technicalities of the index construction, it is clear that to interpret the index in terms of actual real numbers would have demanded information about corruption far beyond any conceivable possibility.

For example, to give any clear meaning of the mathematical *ratio* between two levels of corruption the method of measuring corruption has to be clearly established in beforehand. Imagine that bribes represent 10% of *all* payments in country X. In country Y bribes represent only 2% of *half* of all transactions. If we measure the *volume* of corruption, that is the total sum paid in bribes, then country X is 10 times as corrupt as country Y. However, if we measure the *number* of corrupt transactions, country X is only twice as corrupt as Y. Furthermore, given the unclear measures, we can not say if the difference in corruption levels between two countries with indexes 3 and 4, is identical with the difference in corruption levels between two countries with indexes 5 and 6.

Accordingly, if quantification of corruption is uncertain, comparison of corruption levels among countries is also problematic. The question of the *extent* of corruption is specified in the surveys. However, when it comes to the aggregate indexes, and the CPI in particular, it is a problem that they often are based on a *combination of specific descriptions*. Thus, although the aim is more evident information, this approach might result in *veiled* information. But regardless of an uncertain measure of the “level of corruption”, such “ranking” are informative, since they may provide us with information on whether corruption is “most likely” to be a “more serious” problems in one country compared to others.

Finally, it is also important to underline the partial nature of the CPI. As it was mentioned above, the CPI reflects mainly the perceptions of the international business-community. This informant group along with some international organizations has put corruption back on the international policy agenda. There is nothing wrong with this, and in fact the study of corruption is richer today for the efforts of such groups. However, these observations and opinions are partial visions. To know how corruption –conceptualized as bribery— affects development – expressed in GDP figures or in terms of governance indicators— is valuable knowledge, but there is much more to be said. Years ago, for example, Huntington (1968) argued that corruption might be a preferable alternative to violence. So far, there is not any analysis providing empirical data to reject or accept such an argument. Statistical indices can not settle that sort of question by themselves. We will also need historical, political, and cultural evidence, and knowledge of forms of corruption beyond the bribery paradigm. It is important, for instance, to know about the nature of accountability and justice, the dynamics behind a corrupt cooperation,

the role of normative frameworks and its effects in corrupt scenarios, and so on. No index could be expected to reveal these subtleties, but they are no less important for being less easily quantified.

These are more than methodological details. Perceptions of corruption do shape important decisions, but the danger is that they will become to an “echo chamber” problem in which officials and investors repeat what they hear from each other. Analysts can make good use of perceptions of corruption, but there must also be ways to anchor perceptions in less subjective information about societies.

Corruption Indexes: Second-Generation Measures

It is unlikely that we will ever have valid, reliable, precise, and broadly comparable data on corruption. But even if we had, they would be only one aspect of the broader and richer comparisons that are needed for both analysis and reform. Understanding the different forms of corruption, their links to causes and their consequences, requires many kinds of evidence and theoretical approaches sensitive to variations among societies. Reforms and more general development efforts need similar foundations. A number of attempts have been made to improve our measurements of corruption. However, the real challenge for the next stage of corruption research is not just to improve our measurement, but rather to build a richer understanding of the phenomenon, and to show why such an understanding is essential.

The first-generation indices provoked strong reactions. Journalists present the CPI as a rating of the world’s most corrupt countries, even though Transparency International explicitly warned against that interpretation. International agencies and many scholars quickly put the data to work, while others were more critical.

As a consequence of this discussion a new set of initiatives have been arisen. These initiatives could be named “second-generation” measures. A variety of sample surveys have focused on the *experiences* of corruption rather than perceptions of business, households, and individuals. These surveys are subject to many of the same validity, reliability, and precision problems discussed earlier, but provide a level of detail that first-generation scales can not offer. A variety of organizations, including the US Agency for International Development, have sponsored such surveys. The most elaborate is the World Bank Institute’s *Business Environment and Enterprise Performance Survey (BEEPS)*. This survey was carried out in twenty transitional states in the former USSR and Eastern and Central Europe (Hellman, *et. al* 2000b). The data, based on a seventy-item survey of business firms and on some supplementary questions, were gathered in 1999. The survey approach allows researchers to consider different varieties of corruption. However, this survey presents problems common to other cross-national surveys. These kind of surveys involve obvious linguistic problems. Other difficulties of comparison, such as a tendency for respondents to under or overestimate the corruption with which they deal, must be taken into account too. However, it is also important to note that the BEEPS survey asks respondents’ views on verifiable aspects of the business

environment, such as exchange-rate fluctuations. Doing so, perceptions can be checked against valid indicators, allowing an intelligent guess as to whether extent respondents systematically over or underestimate the level of corruption.

BEEPS-style projects are formidably expensive and, while including a wider variety of corrupt practices and situations than most other indices, still approaching the problem of corruption from the perspective of businessman. Nonetheless, they are an extremely promising addition to the growing number of corruption measures. The BEEPS 1999 data have already begun to produce some comparative studies, which bring important light about corruption in transitional economies (Hellman, *et al.* 2000b).

In 1999, Transparency International published its *Bribe Payers Index (BPI)*. This index ranked nineteen leading exporting countries in terms of their own firms' propensity to pay bribes to public officials. The BPI is based on Gallup survey data gathered in fourteen "emerging market" countries. The results were quite different from those of the CPI. Countries such as Sweden, Australia, Canada, the United Kingdom, and the United States came off much less favorably in this index. However, this index has its own methodological problems too. Again it is the problem of the differences between perceptions and corruption itself: Are respondents basing their judgments on actual knowledge of corruption or on their general opinions of particular countries? The BPI is a new measure and it allows to raise important issues about the sources of bribes. Nevertheless, this index must be also taken just as preliminary method of measurement and not as a final ranking method.

Kaufmann, Kraay and Zoido-Lobaton (1999a, 1999b) have constructed a sophisticated index of "graft" as a part of a broader measure of the quality of governance. Governance, as the concept of corruption, is difficult to define and measure. Kaufmann and his colleagues defined governance in terms of three sub-concepts in an attempt to quantify governance. These sub-concepts are: *a)* probity, *b)* bureaucratic quality, and *c)* rule of law. In order to measure them, the authors use thirty-one component measures that allow the inclusion of 166, 156, and 155 countries, respectively, in aggregate indices of the rule of law, government effectiveness, and graft (Kaufmann, Kraay and Zoido-Lobaton, 1999b). Unlike the CPI, which excludes countries with fewer than three corruption surveys, this approach is more inclusive. The argument behind this index is that data are likely to be more plentiful for countries with better governance, and excluding those for which data are scarce has the effect of omitting many of the worst-governance cases, as discussed before. Supporting this view, Kaufmann and his colleagues (1999c) conducted a comparison between preliminary estimates of "probity" and 1998 CPI scores. The CPI produced systematically lower estimates which are most likely to be the result of excluding countries with fewer than three surveys.

Kaufmann, Kraay and Zoido-Lobaton found that even given some strong assumptions, standard errors for their governance indicators (including graft) were very large. It was possible to identify a handful of countries at the good and bad

governance extremes, but results for the vast majority did not differ statistically from the global means. In other words, for most countries, the data did not support confident judgements that probity, bureaucratic quality, and rule of law were particularly high or low (Kaufmann *et al.*, 1999a: 15-19). However in this index, as in the case of the CPI, we can not be sure that “real” variations are equal across all identical intervals of the indexes. To some extent, the most realistic corruption measure can do is to identify bands of cases likely to have broadly comparable levels of corruption.

An alternative approach (Hall and Yago, 2000) focuses on the concept of “opacity” –the opposite of transparency or, in other words, restrictions upon the open flow of information essential to orderly, efficient markets. Opacity has many forms in practice, ranging from false accounting to intimidation, and serves “to ensure the secrecy of corrupt or questionable practices” (Hall and Yago, 2000: 1). The index is based on a statistical model that incorporates corruption in several different ways. CPI scores are included as an estimate of corruption, along with macroeconomic data and various measures of institutional quality. The data are used to account for the varying interest rates paid by governments as they float bonds on the international market. Those with poorer institutions and higher levels of opacity pay higher costs –a “premium”–to borrow money. Estimating the size and sources of these costs is, in effect, a way to compare the seriousness of these countries’ corruption problems. The authors calculate an “institutions premium”, a “corruption premium”, and “graft premium” –this last indicator based on the graft index calculated by Kaufmann *et al* (1999a).—for each of thirty-five countries. These estimate “the shortfall each country had from the perfect transparency score” (Hall and Yago, 2000: 5). The results are strong and consistent. Poor-quality institutions, corruption, and graft are linked to significantly higher costs of borrowing –estimated at over \$130 billion per year for the sample of thirty-five economies. These are indirect measures of corruption, but they have the virtue of incorporating perceptual scales into a range of harder indicators.

However, while the corruption and graft premium are both consistent with the perceptions informed by CPI –and, in practical terms, are likely influenced by such perceptions, as noted earlier—construct validity, reliability, and precision are augmented by the ways lenders continually evaluate countries’ economic performance. In sum, corruption itself remains difficult to measure, but the notion of building indexes based upon more reliable measures of other variables closely related to corruption is a very promising one.

Conclusions

How much guidance do corruption indexes give reformers? Can those fighting corruption in a society look at CPI scores for evidence of progress, and for guidance in shaping their strategies? No, corruption indexes should not be used as policy evaluation instruments. CPI data do exhibit impressive reliability, but as noted before we still do not know how well corruption indexes track changes in levels of corruption. Perceptions are not corruption itself. Any anti-corruption strategy will likely work better with some aspects of the problem than with others. Therefore, a single-number index will not be able to tell us much about which aspects of the strategy are working and which are not.

What is likely to happen to perception scores for a country that has begun to make progress against corruption? There are several possibilities. Progress will be uneven and thus recognized more quickly by some observers than by others. In that event, CPI scores might change in ways that would be difficult to interpret. Another possibility is that a successful anti-corruption campaign would produce revelations of wrongdoing, convictions, and 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 scenario, effective anti-corruption efforts would likely cause perceptions of worse corruption situation, at least in the short run. (Knack and Keefer 1995). Once again, CPI ratings will not be able to tell us about this vagueness. Surveys, whether on the BEEPS scale or smaller, are probably the best way to evaluate anti-corruption policies. But they are expensive and may not reveal much about progress against the deeper causes of corruption, or why observed trends are occurring.

Can we devise relatively inexpensive measures that are still sensitive to changing levels of corruption, and can give useful guidance to anti-corruption efforts? One way might be to focus less on measuring corruption itself and more on scaling its correlated issues. We have good reasons to think that a variety of conditions and phenomena are closely linked to corruption. Many of these have been measured at a considerable level of validity, reliability, and precision, and in ways that do not reify perceptions and anecdotes as broader trends. Serious corruption is deeply embedded, and causality can be difficult to disentangle. Still, we might construct indexes approximating causes and effects of corruption. Loayza (1996) has employed a similar approach in studying informal economies, a measurement challenge resembling corruption in many respects. On the causes side, we could incorporate measures of major problems giving rise to corruption, like poor-quality institutions, lack of political competition, lack of openness in the economy, inflation, and weak guarantees of civil liberties and property rights. On the effects side, it will be necessary an index which includes factors such as budget composition indicators, statistics on the efficiency of tax collection, extension of “black-market” activities,

trends in aggregate development, indicators of various forms of capital flight, and so on. Both indexes (causes and effects of corruption indexes) could be based on an unobserved-components model, and both could be designed to include different forms of corruption. A focus on specific countries or regions over time would reduce the risk of distortions caused by the differing data available in various countries. The result could be complementary indexes to those now available, yet sensitive to changes and to the deeper causes and effects of corruption.

There are some obvious problems with this approach. Endogeneity and simultaneity make causes and effects of corruption difficult to separate. Are ineffective tax collection or black-market activities results of corruption, or do they create incentives that cause it? This approach, while it might reveal distinctive aspects of corruption in particular societies or regions, would not produce “headline numbers” for broad cross-national comparisons. Moreover, in this approach remains problems of reliability and precision. How, for example, should the components of such indexes be used? Should we use a regression model that predicts CPI or other scores for some initial point in time and weight measures by their statistical power to predict our “effects index”, or changes in it? If so, how should we use the components of the effects scale? There would also be questions of how to report the results. Are annual results extended to decimal places appropriate? Would they raise expectations that can not be met or, because of the long-term nature of basic anti-corruption reforms, lead to disillusionment? Would reporting results in broader intervals create the illusion that nothing is changing?

An important thing that have shaped many first-generation and their uses is the gap between qualitative and quantitative methods in the empirical research on corruption. Usually the implications is that broadly comparative works fall into the qualitative kind of studies –particularly that aimed at neopatrimonialism conceptual framework and broad based explanations. While statistical approaches are basically quantitative empirical research.¹⁰ This distinction has provoked serious problems for understanding well the corruption phenomenon. In the current research on corruption there is an overemphasis upon the narrow range of factors of corruption. Perhaps this overemphasis is clearly illustrated by the ways corruption indexes reduce complex cases to single numbers, and encourage cross-sectional statistical approaches that impose a single model on widely divergent cases. Statistical studies on corruption, based on corruption indexes, may lose sight of the historical origins of corruption, and thus of some of the forces and conflicts sustaining it. Cultural and linguistic factors are also important shaping the social significance of corruption, and many of this factors are underestimated by most of the corruption indexes. Therefore, it is important to underline the necessity of more comprehensive methods of measurement and analysis.

There is not valid reason why the qualitative/quantitative bifurcation should exist. The empirical research of corruption needs qualitative sorts of evidence as

¹⁰ For a discussion about the qualitative and quantitative studies on corruption see Del Castillo (2002).

well as quantitative frames of analysis and measurement. Broad-based comparative frameworks merging quantitative evidence with qualitative knowledge, and with linguistic, cultural, and historical evidence, would serve for both reforms and scholar analysis. Reform and analysis will always remain distinct enterprises, but as Hall and Yago's work on "opacity" suggests, the shortcomings of perceptual measures become less critical the more they are augmented with other evidence. Comparative frameworks may generate more precise hypotheses to the extent that they draw upon quantitative and qualitative evidence. Second-generation corruption measures show that as they become more elaborate models they are increasingly distant from "corruption rankings". There is no reason why this trend can not be carried out further. In the future, corruption indexes should be based on questions that include linguistic and cultural factors. They should be indexes that combine qualitative and historical depth, along with the kind of breath that cross-sectional statistical data can provide.

There is no doubt that the effort to measure corruption has been worthwhile. It has helped set to rest a variety of questions that had long kept the scholars debate going around in circles. Even though current measures of corruption still in a preliminary stage and the corruption debate is based on a narrow bribery paradigm. The potential of any research to produce rich and useful insights depends fundamentally on careful design and honest application, not on apparent simplicity of its methods or results. The task now is to bring evidence of many sorts and build more comprehensive indexes. Doing so, it will be possible to carry out more accurate comparative analysis. It is important not to forget that a better understanding of the local realities is essential for an accurate knowledge of this phenomenon and the design of proper anti-corruption policies.

Bibliography

- Andvig, J. and K. Moene (1990) "How Corruption May Corrupt", *Journal of Economic Behaviour and Organization*, 3(1): 63-76.
- Babbie, E. (1995) *The Practice of Social Research*, California, Wadsworth.
- Braun, M. and R. Di Tella (2000) "Inflation and Corruption", *Harvard Business School Working Paper 00-053*, Cambridge, Harvard Business School, Division of Research.
- Collier, D. and S. Levitsky (1997) "Democracy with Adjectives: Conceptual Innovation in Comparative Research", *World Politics*, 49(April): 430-451.
- Del Castillo, A. (2001) "El Soborno: Un Marco Conceptual para su Análisis", *Gestión y Política Pública*, X(2): 375-402.
- Del Castillo, A. (2002) "Let's Talk About Corruption. Qualitative and Quantitative Methods in the Research of Corruption", *Six Essays on Corruption. The Empirical Foundations of Recent Research on Corruption*, Ph.D. Thesis, Department of Administration and Organization Theory, University of Bergen.
- Easterly, W. and R. Levine (1996) "Africa's Growth Tragedy: Policies and Ethnic Divisions", Washington, World Bank, Policy Research Department, Macroeconomics and Growth Division.
- Genaux, M. (2000) "Early Modern Corruption in English and French Fields of Vision", in A. Heidenheimer and M. Johnston (eds.), *Political Corruption*, New Jersey, Transaction.
- Goldsmith, A. (1999) "Slapping the Grasping Hand: Correlates of Political Corruption in Emerging Markets", *American Journal of Economics and Sociology*, 58(4): 866-883.
- Glynn, P., S. Kobrin and M. Naim (1997) "The Globalization of Corruption", in K. Elliott (ed.), *Corruption and the Global Economy*, Washington, Institute for International Economics.
- Hall, T. and G. Yago (2000) "Estimating the Cost of Opacity Using Sovereign Bond Spreads", Working Paper, Santa Monica, Capital Studies Group, Milken Institute.
- Hellman, J., G. Jones and D. Kaufmann (2000a) "Seize the State, Seize the Day. State Capture, Corruption and Influence in Transition", *World Bank Policy Research Paper 2444*.
- Hellman, J., G. Jones, D. Kaufmann, and M. Schankerman (2000b) "Measuring Governance, Corruption, and State Capture: How Firms and Bureaucrats Shape the Business Environment in Transition Economies", *World Bank Policy Research Working Paper 2312*.
- Huntington, S. (1968) *Political Order in Changing Societies*, New Haven, Yale University Press.
- Johnston, M. (2000) "Party Systems, Competition, and Political Checks against Corruption", in A. Heidenheimer, M. Johnston (eds.), *Political Corruption*, New Jersey, Transaction.
- Kaufmann, D., A. Kray and P. Zoido-Lobaton (1999a) "Aggregating Governance Indicators", *World Bank Policy Research Paper*, n. 2195, Washington DC, World Bank.
- Kaufmann, D., A. Kray and P. Zoido-Lobaton (1999b) "Governance Matters", *World Bank Policy Research Paper*, n. 2196, Washington DC, World Bank.
- Kaufmann, D., A. Kray and P. Zoido-Lobaton (1999c) "Aggregating Governance Indicators", (August), Washington DC, World Bank, unpublished.
- Knack, S. and P. Keefer (1995) "Institutions and Economic Performance: Cross-Country Test Using Alternative Institutional Measures", *Economics and Politics*, 7(3): 207-227.

- Lambsdorff, J. (1999a) "Corruption in International Research –A Review", *Transparency International Working Paper*, Berlin, www.transparency.de [10.11.00].
- Lambsdorff, J. (1999b) "The Transparency International Corruption Perception Index 1999 – Framework Document", *Transparency International Working Paper*, Berlin, www.transparency.de [13.12.00].
- Lancaster, T. and G. Motinola (1997) "Toward a Methodology for Comparative Study of Political Corruption", *Crime, Law and Social Change*, 27(3-4), p.p. 185-206.
- Loayza, N. (1996) "The Economics of the Informal Sector: A Simple Model and Some Empirical Evidence from Latin America", *Carnegie Rochester Conference Series on Public Policy*, 45: 129-162.
- Mauro, P. (1995) "Corruption and growth", *Quarterly Journal of Economics*, CX(3): 681-712.
- Mauro, P. (1997) "The Effects of Corruption on Growth, Investment, and Government Expenditure: A Cross-Country Analysis", *Corruption and the Global Economy*, Washington, Institute for International Economics, p.p. 83-107.
- Moroff, H. and V. Blechinger (2000) "Corruption Terms in the World Press: How Languages Differ", in A. Heidenheimer and M. Johnston (eds.), *Political Corruption*, New Jersey, Transaction.
- Ocampo, L. (1993) "En defensa propia: ¿cómo salir de la corrupción?", *Suramericana*, 5(4): 20-25.
- Rose-Ackerman, S. (1999) *Corruption and Government. Causes, Consequences and Reform*, Cambridge, Cambridge University Press.
- Schelesinger, T. and K. Meier (2000) "Variations in Corruption among the American States", in A. Heidenheimer and M. Johnston (eds.), *Political Corruption*, New Jersey, Transaction.
- TILAC (1999) "Another Form to Look at the TI Index: Latin America as an Example", unpublished.
- United Nations (1999) *Global Report on Crime and Justice*, Oxford, Oxford University Press.
- Wei, S. (1997) "How Taxing Is Corruption on International Investors?" *National Bureau of Economic Research Working Paper 6030*, Cambridge, National Bureau of Economic Research.