**Political Concepts**
Committee on Concepts and Methods
Working Paper Series

48

March 2011

# Does the Evaluator Make a Difference?

## Measurement Validity in Corruption Research

Angela Hawken

Pepperdine University (angela.hawken@pepperdine.edu)

&

Gerardo L. Munck

University of Southern California (munck@usc.edu)

**C&M**
The Committee on Concepts and Methods
www.concepts-methods.org

**IPSA**
International Political Science Association
www.ipsa.org

**CIDE**
Research and Teaching in the Social Sciences
www.cide.edu

**The C&M working paper series** are published by the Committee on Concepts and Methods (C&M), the Research Committee No. 1 of the International Political Science Association (IPSA), hosted at CIDE in Mexico City. C&M working papers are meant to share work in progress in a timely way before formal publication. Authors bear full responsibility for the content of their contributions. 

**The Committee on Concepts and Methods (C&M)** promotes conceptual and methodological discussion in political science. It provides a forum of debate between methodological schools who otherwise tend to conduct their deliberations at separate tables. It publishes two series of working papers: "Political Concepts" and "Political Methodology."

**Political Concepts** contains work of excellence on political concepts and political language. It seeks to include innovative contributions to concept analysis, language usage, concept operationalization, and measurement.

**Political Methodology** contains work of excellence on methods and methodology in the study of politics. It invites innovative work on fundamental questions of research design, the construction and evaluation of empirical evidence, theory building and theory testing. The series welcomes, and hopes to foster, contributions that cut across conventional methodological divides, as between quantitative and qualitative methods, or between interpretative and observational approaches.

**Submissions**. All papers are subject to review by either a member of the Editorial Board or an external reviewer. Only English-language papers can be admitted. Authors interested in including their work in the C&M Series may send their paper to wp[at]concepts-methods.org. They may also seek initial endorsement by one editorial board member.

The C&M webpage offers full access to past working papers.

**www.concepts-methods.org**

* Jeremy Grunert, Michelle Isenhouer, Matt Leighty, Matthew C. Piccolo and Matthew Turvey assisted us in the preparation of this paper. Their work is greatly appreciated. We also thank Aart Kraay for his comments.

The nature of corruption, as well as its causes and consequences, has been explored in many empirical studies over the past four decades (Heidenheimer and Johnston 2002). However, an important new element was added to this literature as a result of the production of quantitative corruption data spanning most countries of the world. These data were initially accessible only from commercial risk assessors. But the subsequent incorporation of this data, as well as data gathered from other sources, into two publicly available corruption indices—Transparency International's Corruption Perception Index (CPI), published since 1995, and the World Bank's Control of Corruption Index (CCI), available since 1999—greatly expanded access to quantitative data on corruption. And the gradual incorporation of these measures into academic research led to a new quantitative, cross-national literature on corruption (Lambsdorff 2006a, Treisman 2007).

This paper focuses on the quality of the data used in the corruption literature. It seeks to contribute to an understanding of available measures of corruption through an analysis of key methodological decisions in the production of indicators of corruption, understood here as simple measures that are commonly used to construct indices, and the production of indices of corruption, that is, compound measures that aggregate multiple indicators. In the first section, we focus on indicators of corruption and conduct a simple test to ascertain whether differences in terms of *who* evaluates a country's level of corruption affect the resulting measures. The result of this test shows that data relying on different classes of indicators, a distinction based on the nature of the evaluator, are systematically associated with higher or lower corruption scores and, moreover, that the differences among classes of indicators also vary across regions of the world. That is, our analysis shows that current measures of corruption do not provide a basis for ascertaining the level of corruption of countries around the globe with much confidence and also uncovers a rather ominous problem: Indicators of corruption, as a group, fail a basic precept of scientific research, that the results of research—in this case research focused on measurement—do not depend upon who conducts the research.

A standard response to questions regarding the validity of existing indicators of corruption has been that any problems concerning indicators can essentially be sidestepped by combining multiple indicators. Indeed, this has been the argument explicitly made by index developers. Thus, in the second section, we focus on the two main indices of corruption, Transparency International's CPI and the World Bank's CCI, and consider whether these indices overcome the problem revealed in the analysis of indicators. We start by addressing how index developers tackle the choice of indicators to include in their indices. We show that the distribution of indicators is not balanced across the available classes of indicators, both over time and across regions, and that the problem with indicators is thus imported into these indices. Thereafter, we turn to the choice of aggregation procedure used to construct indices, that is, the manner in which the values of multiple indicators are combined into a single value, and we consider whether the indicator problem is solved at this stage in the index construction through the formal weighting of indicators. We find that, due to the choice of indicators and formal weighting scheme, the CCI and the CPI unjustifiably—and perhaps unwittingly—assign different weights to different classes of indicators and, furthermore, do not use consistent weighting schemes to measure all countries. And we demonstrate how the choice of weighting scheme induces bias in, and undermines the comparability of, the CCI and CPI data. In short, we show how the development of indices has not solved the problem we uncover at the level of indicators.

This analysis of corruption indicators and indices is unique in the sense that we use data on corruption indicators that have not previously been publicly scrutinized. Indeed, though the quantitative, cross-national literature on corruption has used, directly or more commonly indirectly (via indices), dozens of indicators over the past fifteen years, data on indicators—as opposed to indices—of corruption have not been widely available for academic scrutiny. Thus, even though a growing number of publications and papers has addressed the validity of various measures of corruption (Arndt and Oman 2006, Sampford et al. 2006, Knack 2007),[1] these assessments have been largely conceptual and have not presented empirical tests to substantiate their usually critical view of the most-commonly used measures of corruption. In turn, those who have had privileged access to the full range of corruption indicators—the developers of the CPI and the CCI—have failed to appropriately test the validity of their indices or offer a useful assessment of existing indicators of corruption.[2] Thus, this paper presents the first tests using the

[1] See also Kurtz and Schrank (2007a, 2007b). Iqbal and Shah (2008), Langbein and Knack (2008), Mehlkop, Graeff, and Neumann (2008), and Thomas (2010).

[2] The developers of the CPI and the CCI have largely presented the results of tests to defend choices that had been made previously, rather than a way to assess the theory guiding the construction of the index. For example, tests on the robustness of the data to changes in the aggregation rule were apparently performed well after the design of the indices. Though the CPI was released in 1995, the robustness test reported in 1999 seem to have been conducted in response to the appearance of the CCI (Lambsdorff 1999: 18). In turn, though the World Bank's CCI was introduced in 1999, it appears that it was not until 2006 that a robustness test was conducted (Kaufmann, Kraay, and Mastruzzi

full range of data sets used in cross-national corruption research to shed light on the critical issue of the validity of corruption data (on our data, see Appendix 1).

We conclude that the problems with current corruption data are serious enough to call for a change in standard practices. In a nutshell, corruption data suffer from a fundamental problem: Different data sets used in quantitative research are routinely associated with different findings, and the relative validity of different measures of corruption and hence of the different findings is not readily apparent. Thus, after highlighting the problems with available corruption data, we offer some thoughts about improving the generation of corruption data and their use in the quantitative, cross-national literature on corruption.

## I. Classes of Corruption Indicators

The quantitative, cross-national literature on corruption has been made possible by the production of a variety of corruption indicators—simple measures in contrast to compound measures that aggregate multiple indicators. The number of available indicators is high; for the purpose of the empirical analysis in this section, we draw on 23 distinct indicators that cover 210 countries and territories over nine years (1996, 1998, 2000, and 2002–07). These indicators have

---

2006: 25). Indeed, the results of tests presented by the developers of the CPI and the CCI seem like an afterthought rather than part of an open-ended inquiry about the quality of corruption data. And, most fundamentally, a review of all the methodological papers written by these index developers, starting with the first discussion of the CPI in 1995 and of the CCI in 1999, shows that the tests that were performed are quite weak, frequently confusing the concepts of validity and reliability.

Index developers have reported correlations among individual sources of data on the indicators that serve as underlying sources for their indices (Transparency International and Lambsdorff 1997: 11-12; Lambsdorff 2006c: 8, 2007: 248-50; Kaufmann, Kraay and Mastruzzi 1999: 31, 2006: 42). But they have not capitalized on the opportunity to apply standard methods to detect the presence of bias, and to determine the factors driving any such bias. Rather, inasmuch as tests of a possible evaluator bias have been performed, these tests rely on patently implausible assumptions, such as that one can simply designate a measure as the correct measure of corruption (Kaufmann, Kraay and Mastruzzi 2004: 273-75), or on basic mistakes, such as passing a means to estimate reliability for one to address the core matter of validity and relying solely on correlations among indicators (Kaufmann, Kraay and Mastruzzi 2006: 20-25, 41; 2007a: 556-57; Lambsdorff 2006b: 86-87, 2008: 6).

Likewise, index developers have presented the results of some tests directly pertaining to the construction of indices, such as a test of the independence of sources (Kaufmann, Kraay and Mastruzzi 2006: 25-31, 42-44; Lambsdorff 2006c: 6) and of the robustness of indices to changes in aggregation rule (Lambsdorff 1999: 18, Kaufmann, Kraay and Mastruzzi 2006: 25, 2007c: 21-22). But the impact of the weights assigned to indicators was tested on an unrepresentative sample and the possible impact of variations in the number of underlying sources was tested in a flawed manner, on a small set of cases selected on the dependent variable and that by design addresses the impact of only a few sources that have a low weight in the index (Kaufmann, Kraay and Mastruzzi 2005: 12-14, 55-57; 2007d: 20-21, 34; 2008: 22, 35). In short, though Transparency International and the World Bank have explained in some detail how they constructed their respective indices, the CPI and CCI, they have fallen short of providing a thorough validation of the data on the indicators of corruption they use as a basic input in their indices and the methodological choices that go into the making of their indices.

been used regularly in corruption research since the mid-1990s, either directly or indirectly, when these indicators are used in indices. But they have not been analyzed and/or compared empirically in much depth. Thus, we start our evaluation of corruption data by considering a set of indicators that are routinely interpreted as corruption measures.

One central methodological difference in the way that data on indicators are generated concerns *who* evaluates a country's level of corruption. After all, the indicators we consider rely on responses to questions by evaluators and, using characteristics of the evaluator as the criterion of classification, it is possible to distinguish among five classes of sources of data on corruption indicators—classes of indicators, for short: Those that rely on i) expert ratings by a commercial risk assessment agency, ii) expert ratings by an NGO, iii) expert ratings by a multilateral development bank (MDB), iv) surveys of business executives, and v) surveys of the mass public (for a classification of indicators, see the note accompanying Table 1). It is important therefore to test whether the results of the measurement process are unrelated to the characteristics of the evaluator, a basic principle of scientific measurement, or, alternatively, whether there are grounds for arguing that a variable feature of the measurement instruments—their reliance on different evaluators—as opposed to actual differences in corruption, affects measures of corruption.[3]

As a first step to see if the choice of evaluator has an impact on the resulting measures of corruption, we conducted a comparison of global means of pairs of cases (country-year) measured with two classes of indicators. This is a basic test that compares how two classes of indicators describe the same set of cases. And the results support a strong conclusion (see Table 1): Different classes of indicators yield measures that vary significantly in terms of their strictness, that is, whether they tend to systematically generate higher or lower estimates of the level of corruption in countries around the world. Specifically, only the data generated by expert ratings by MDBs and surveys of the mass public are indistinguishable.

[3] The literature on this question has been ambiguous. Though some authors have played down the divergence among measures of corruption produced by different classes of indicators (Treisman 2000: 412, Kaufmann, Kraay and Mastruzzi 2007a: 556-57), others have noted the difference between classes of indicators, emphasizing in particularly the contrast between common citizens and other respondents (Razafindrakoto and Roubaud 2006: 21-23, Kurtz and Schrank 2007b: 566).

**Table 1. Classes of Indicators I: Difference Between Means, Global Comparisons**

| **Order of Stricness** (">" is "stricter than") | **Difference between Means** (0.000-1.000 scale) | **Significance** (p value) | **Standard Deviation** |
|---|---|---|---|
| Expert Commercial > Business Survey | 0.059 | 0.000 | 0.165 |
| Expert Commercial > Expert NGO | 0.079 | 0.000 | 0.128 |
| Expert Commercial > Mass Survey | 0.146 | 0.000 | 0.219 |
| Expert Commercial > Expert MDB | 0.147 | 0.000 | 0.151 |
| Expert NGO > Mass Survey | 0.029 | 0.007 | 0.195 |
| Expert NGO > Business Survey | 0.031 | 0.005 | 0.228 |
| Expert NGO > Expert MDB | 0.088 | 0.000 | 0.118 |
| Business Survey > Mass Survey | 0.073 | 0.000 | 0.195 |
| Business Survey > Expert MDB | 0.071 | 0.000 | 0.172 |
| Expert MDB > Mass Survey | 0.005 | 0.605 | 0.165 |

*Note:* MDB = multilateral development bank. The difference between means is calculated through paired comparisons of all countries for which indicators from the two classes of sources are available (when more than one indicator of a certain class is available, as is usually the case, the values of these indicators are averaged). The data cover 210 countries over the 1996, 1998, 2000, and 2002-07 years.

The Expert Commercial class includes the following data sets: the Business Environment Risk Intelligence's Business Risk Service, the Business Environment Risk Intelligence's Qualitative Risk Measure in Foreign Lending, the Economist Intelligence Unit's Country Risk Service, Global Insight's Business Conditions and Risk Indicators, Global Insight's Global Risk Service, the Merchant International Group's Grey Area Dynamics, and the Political Risk Services' International Country Risk Guide.

The Expert NGO class includes the Bertelsmann's corruption measure in the Bertelsman Transformation Index, Freedom House's corruption measures in *Countries at the Crossroads* and *Nations in Transition*.

The Expert MDB class includes the African Development Bank's Country Policy & Institutional Assessment (CPIA), the Asian Development Bank's CPIA, the IFAD's Rural Sector Performance Assessments, and the World Bank's CPIA.

The Business Survey class includes the Institute for Management Development's Executive Opinion Survey in the *World Competitiveness Yearbook*, the Political Economic Risk Consultancy's Corruption Survey, the World Bank's Business Enterprise Environment Survey, and the World Economic Forum's Global Competitiveness Survey in the *Global Competitiveness Report*.

The Mass Survey class includes the Afrobarometer Survey, Latinobarometro Survey, Vanderbilt University Americas Barometer Survey, Transparency International's Global Corruption Barometer Survey, and Gallup's World Poll.

*Data Source:* Calculations based on a data set constructed from World Bank (2009a); see Appendix 1 for more detail.

But the finding is even stronger. The relationship among the five classes of indicators follows a highly structured pattern, whereby each class of indicator can be ordered, in descending order of strictness, as follows:

*Expert Commercial > Expert NGO > Business Survey > Expert MDB = Mass Survey*

This is a striking pattern, which does not seem consistent with random measurement error. This pattern is evidence that a country's assessed level of corruption is affected by who conducts the assessment, as some evaluators are distinguishably stricter or more lenient than others. Different evaluators have, in general, different standards.

To gain a sense of the magnitude of the difference that can be interpreted as systematic measurement error, it is useful to compare the level of variability that is attributable to these different classes of evaluators as opposed to subjective measures in general, that is, measures that rely on judgments made by people in response to certain questions. It could very well be the case that all subjective measures are highly unreliable, and hence that these differences—though indicating a problem in the measurement of corruption—are suggestive of random error. Yet the evidence is not encouraging. A reasonable estimate of the level of variation that can be expected from subjective measures of corruption is between 1.5 and 3.0% of the overall variation in an indicator scale.[4] In contrast, the differences in these data on corruption, again relative to the overall variation in the indicator scales, reaches as high as 14.7% in the case of the global comparison between expert ratings by commercial risk assessment agencies and the expert ratings by MDBs (see Table 1). The differences in the corruption scores of different classes of indicators are not minor.

We next considered whether the difference in standards across evaluators hold consistently across countries. We compared the regional means relative to the global mean of each class of indicator, again calculating these mean values using paired comparisons of all countries for which data from the two classes of indicators are available, and sought to ascertain whether any class of indicator assesses any region according to stricter or more lenient standards than it treats all cases in the world on average. And, again, the results of the empirical analysis were revealing (see Table 2).

---

[4] This baseline figure was calculated by comparing the mean values on 16 governance variables (one of them being corruption) for 2005 that were coded by expert raters of three multilateral development banks (the World Bank, the African Development Bank, and the Asian Development Bank) using a common set of guidelines (the CPIA framework). For the framework, see World Bank (2005); the data are drawn from ADB (2006: 13), AfDB (2006: Table 1), and World Bank (2006). Through this comparison, we are able to hold the class of respondent constant and hence assess the range of variability that might be due solely to the fact that measures are assigned by experts, as well as to consider whether subjective measures of corruption are different from subjective measures of other governance indicators. The overall difference between the mean of the scores assigned both by the World Bank and the African Development Bank on the 15 variables (excluding corruption) was 1.3% ($p = 0.000$, $N = 585$), with the World Bank assigning the lower scores on average. The difference between scores assigned both by the World Bank and the Asian Development Bank on the 16 variables was 2.2% ($p = 0.000$, $N = 285$), the lower mean corresponding again to the World Bank. The overall difference between the mean of the scores of the corruption variable was 3.0% ($p = 0.032$, $N = 39$) when the coding was done by the World Bank and the African Development Bank, and 2.2% ($p = 0.135$, $N = 19$) when the coding was done by the World Bank and the Asian Development Bank. Thus, it seems that a band of variability between 1.5 and 3.0% in mean values might be expected from subjective measures.

**Table 2. Classes of Indicators II: Differences Between Means, Regional Comparisons Relative to Global Comparisons**

**Panel a. Expert Commercial Class**

| Region | Expert Commercial vs. Business Survey | Expert Commercial vs. Expert NGO | Expert Commercial vs. Mass Survey | Expert Commercial vs. Expert MDB |
|---|---|---|---|---|
| Australasia | 0.067 | .. | 0.308 ** | .. |
| North America | 0.153 *** | .. | 0.330 *** | .. |
| Western Europe | 0.142 *** | .. | 0.219 *** | .. |
| Caribbean | -0.043 *** | -0.116 *** | 0.280* | 0.129 |
| MENA | -0.052 *** | 0.143 *** | 0.106 | -0.023 *** |
| South Asia | 0.035 | -0.033 *** | -0.001 *** | 0.022 *** |
| Eastern Europe & Baltics | -0.007 *** | -0.033 *** | 0.119 | 0.000 *** |
| Sub-saharan Africa | -0.034 *** | -0.019 *** | -0.081 *** | 0.017 *** |
| East Asia | 0.091 ** | -0.113 *** | 0.149 | -0.101 *** |
| S.E. Asia | -0.054 | -0.004 *** | -0.066 *** | -0.066 *** |
| Latin America | -0.009 *** | -0.091 *** | -0.154 *** | -0.122 *** |
| Post-Soviet | -0.316 *** | 0.087 | -0.075 *** | -0.022 *** |

**Panel b. Expert NGO Class**

| Region | Expert NGO vs. Mass Survey | Expert NGO vs. Business Survey | Expert NGO vs. Expert MDB | Expert NGO vs. Expert Commercial |
|---|---|---|---|---|
| Caribbean | 0.222** | 0.097 | 0.092 | 0.116 *** |
| East Asia | 0.291 *** | 0.244 *** | -0.036 ** | 0.113 *** |
| South Asia | 0.008 | 0.120 *** | 0.043 *** | 0.033 *** |
| Eastern Europe & Baltics | 0.112 *** | 0.082 *** | 0.030 ** | 0.033 *** |
| Latin America | -0.037 ** | 0.113 *** | -0.010 *** | 0.091 *** |
| Sub-saharan Africa | -0.022 *** | 0.085 *** | 0.031 *** | 0.019 *** |
| S.E. Asia | -0.079 *** | 0.111 *** | -0.030 *** | 0.004 *** |
| MENA | 0.024 | -0.142 *** | -0.095 *** | -0.143 *** |
| Post-Soviet | -0.140 *** | -0.344 *** | -0.048 *** | -0.087 |

**Panel c. Expert MDB Class**

| Region | Expert MDB vs. Mass Survey | Expert MDB vs. Expert Commercial | Expert MDB vs. Expert NGO | Expert MDB vs. Business Survey |
|---|---|---|---|---|
| East Asia | .. | 0.101 *** | 0.036 ** | 0.136 *** |
| MENA | 0.156 *** | 0.023 *** | 0.095 *** | -0.040 |
| Latin America | 0.031 | 0.122 *** | 0.010 *** | 0.091 *** |
| S.E. Asia | -0.067 * | 0.066 *** | 0.030 *** | 0.120 *** |
| Caribbean | 0.133 | -0.129 | -0.092 | 0.017 *** |
| South Asia | 0.010 | -0.022 *** | -0.043 *** | 0.050 *** |
| Eastern Europe & Baltics | 0.023 | 0.000 *** | -0.030 ** | -0.050 |
| Sub-saharan Africa | -0.040 *** | -0.017 *** | -0.031 *** | 0.002 *** |
| Post-Soviet | -0.058 | 0.022 *** | 0.048 *** | -0.301 *** |

**Panel d. Business Survey Class**

| Region | Business Survey vs. Mass Survey | Business Survey vs. Expert MDB | Business Survey vs. Expert Commercial | Business Survey vs. Expert NGO |
|---|---|---|---|---|
| Post-Soviet | 0.201 *** | 0.301 *** | 0.316 *** | 0.344 *** |
| Australasia | 0.230 *** | .. | - 0.067 | .. |
| MENA | 0.107 | 0.040 | 0.052 *** | 0.142 *** |
| Eastern Europe | 0.142 *** | 0.050 | 0.007 *** | - 0.082 *** |
| Caribbean | 0.327*** | -0.017 *** | 0.043 *** | -0.097 |
| North America | 0.138 * | .. | -0.153 *** | .. |
| Sub-saharan Africa | -0.053 *** | -0.002 *** | 0.034 *** | -0.085 *** |
| Latin America | -0.142 *** | -0.091 *** | 0.009 *** | -0.113 *** |
| South Asia | 0.022 | -0.050 *** | -0.035 | -0.120 *** |
| S.E. Asia | -0.127 *** | -0.120 *** | -0.054 | -0.111 *** |
| Western Europe | 0.082 | .. | -0.142 *** | .. |
| East Asia | 0.048 | -0.136 *** | -0.091 ** | -0.244 *** |

**Panel e. Mass Survey Class**

| Region | Mass Survey vs. Expert Commercial | Mass Survey vs. Expert NGO | Mass Survey vs. Business Survey | Mass Survey vs. Expert MDB |
|---|---|---|---|---|
| S.E. Asia | 0.066 *** | 0.079 *** | 0.127 *** | 0.067 * |
| Latin America | 0.154 *** | -0.080 ** | 0.145 *** | -0.037 * |
| Sub-saharan Africa | 0.081 *** | 0.022 *** | 0.053 *** | 0.040 *** |
| Post-Soviet | 0.075 *** | 0.140 *** | -0.201 *** | 0.058 |
| South Asia | 0.001 *** | -0.008 | -0.022 | -0.010 |
| Eastern Europe & Baltics | -0.119 | -0.112 *** | -0.142 *** | -0.023 |
| MENA | -0.136 | -0.024 | -0.107 | -0.156*** |
| Western Europe | -0.219 *** | .. | -0.082 | .. |
| North America | -0.330 *** | .. | -0.138 * | .. |
| Australasia | -0.308 ** | .. | -0.230 *** | .. |
| Caribbean | -0.280 * | -0.222 ** | -0327 *** | -0.133 |
| East Asia | -0.149 | -0.291 *** | -0.048 | .. |

*Note:* The figures are the difference between means (calculated through paired comparisons of all countries for which two classes of indicators are available) for a region relative to the difference between mean for the entire world. Thus, in the Expert Commercial vs. Business Survey comparison, since the mean difference between these sources for Oceania is 0.008 and the mean difference between these sources for the entire world is -0.059, the relative mean difference for this region is 0.067. Positive figures indicate that the first source (e.g. the Expert Commercial in the Expert Commercial vs. Business Survey comparison) offers a more lenient assessment in a certain region relative to its overall degree of leniency/toughness, while a negative figure is indicative of a tougher relative assessment. The regions are presented in descending order, from those most favored to those most disfavored by a class of iindicator. The regions above the line are favored, on average, across the four possible comparions; those below are disfavored. The data cover 210 countries over the 1996, 1998, 2000, and 2002-07 years. Significance levels = * 10%, ** 5%; *** 1%. .. = no observations. If there are no observations for all four comparisons for a region, that region is not included in the table. MDB = multilateral development bank. For the list of countries included in each region, see the Appendix 2.

*Data Source:* Calculations based on a data set constructed from World Bank (2009a); see Appendix 1 for more details. For the list of sources included under each class of indicator, see the note in Table 3

The regional comparisons are quite complex. No single overriding pattern emerges, in the sense that a given region is consistently penalized or favored by all classes of indicators (though the post-Soviet region and East Asia do stand out in this regard). But within each class of indicator, there is some evidence that the standards applied by evaluators differ from region to region. Specifically, expert ratings by commercial risk assessment agencies are relatively favorable in their assessments of countries in Australasia (Australia and New Zealand), North America (USA and Canada) and Western Europe, on the one hand, and relatively unfavorable in their assessments of post-Soviet countries (Russia and other ex-USSR countries), Latin America, and South East Asia. Expert ratings by NGOs favor the Caribbean, South Asia, and Eastern Europe and the Baltics across the board and disfavor post-Soviet countries. And expert ratings by MDBs favor East Asia and Latin America but do not disfavor any region across the board. In contrast to ratings by experts of different sorts, surveys of business executives favor post-Soviet countries and the countries of the Middle East and North Africa (MENA) and disfavor South Asia, South East Asia, and East Asia. And, again in contrast to ratings by experts, surveys of the mass public favor South East Asia and sub-Saharan Africa, and, disfavor East Asia, the Caribbean, Australasia, North America, and Western Europe. In short, different evaluators show considerable differences in terms of how they assess different regions of the world.[5]

Recapitulating, the answer to the question, Does the evaluator make a difference? must be answered affirmatively. As the analysis of indicators shows, a substantial amount of the 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 artifact of the method selected to measure corruption. Different evaluators use different standards in evaluating levels of corruption and, complicating matters further, the difference in standards across evaluators does not hold consistently across countries.[6] As a result, if someone was displeased with the way some indicator depicted the level of corruption in a certain country, this person could search among the many existing indicators until they found an indicator that offered a measure more to their liking. Moreover, the basis for stating that disagreements among evaluators can be adjudicated, by arguing that some indicators are more valid than others and

[5] A comparison of classes of indicators over time also reveals substantial changes in the difference between mean scores of different classes of indicators.

[6] For works that underline and explore the possible biases of respondents, see Lall (2001: 1515-19), Van De Walle (2006: 441-42), Kurtz and Schrank (2007a: 542-43), and Iqbal and Shah (2008: 20-25). On the impact of possible sampling biases, within the context of each source, see Kurtz and Schrank (2007a: 542).

hence that the conclusions drawn from these indicators are more credible, is far from apparent. An inescapable conclusion is that the measurement of corruption includes an element of arbitrariness.

## II. The Weighting of Classes of Indicators in Corruption Indices

One response to concerns about the quality of indicators on corruption has been the construction of indices that combine many of these indicators. The argument, briefly, is that, by combining the indicators, the measurement error of the individual indicators is reduced.[7] Whatever the problems with indicators on corruption might be, at least it would follow that indices on corruption would clearly be preferable and should be given greater credibility compared to indicators on corruption. But the cancelling out of errors does not follow automatically simply because individual indicators are combined. This benefit hinges on the assumptions that the error in the indicators is random as opposed to systematic, and independent across sources. And if these assumptions do not hold, the problems with individual indicators could simply be carried over to an index or even exacerbated, if the true signal in the available indicators is muted because the indicators with the greatest error are given greater weight.

To assess the quality of corruption indices, we evaluate two key choices involved in index construction: i) the choice of indicators (i.e. what indicators are combined in an index) and ii) the choice of aggregation procedure (i.e. the manner in which the values of all the indicators are combined into one single value). In short, our analysis examines the two main indices of corruption used in the literature—the World Bank's Control of Corruption Index (CCI) and Transparency International's Corruption Perception Index (CPI)—and considers whether the developers of these indices address the problems with indicators of corruption through their choice of indicators and aggregation procedure.

### II. i. The Choice of Indicators

Both the CCI and the CPI are constructed on the basis of a large number of sources that provide data on indicators. In recent years, the CCI has been based on approximately two dozen such sources (Kaufmann, Kraay and Mastruzzi 2008: 78), while the CPI has been based on

[7] This argument is made, unsurprisingly, by the developers of indices. For example, the developers of the CCI state that "aggregate indicators can provide more precise measures of governance than individual indicators" (Kaufmann, Kraay and Zoido-Lobatón 2002: 6; see also Kaufmann, Kraay and Mastruzzi 2006: 7, 12)). On the rationale offered for the CPI, see Lambsdorff (2006b: 91; 2006c: 10).

roughly one dozen (Lambsdorff 2008: 2, 12-13). In general terms, the criterion for selection of data sources used by the developers of the CCI is to rely on as many data sources as possible (Kaufmann, Kraay and Mastruzzi 2008: 4), even when they might not be measures of corruption per se; in contrast, the developer of the CPI is more selective, paying more attention to whether the sources actually measure corruption as opposed to some other, potentially associated concept (Lambsdorff 2007: 238-40). Nonetheless, a major concern for the developers of both the CCI and the CPI is to cover most countries of the world and to rely on multiple sources on each country,[8] and for this reason the selection of data sources is driven largely by considerations of convenience, that is, whether data sources covering different countries are available.

The consequences of this sampling of data sources largely by convenience are clear. As the distribution of the classes of indicators over time (see Table 3) and across regions of the world (see Table 4) shows, there is a striking lack of balance in the proportion of different classes of indicators used in both the CCI and the CPI. First, there is considerable variation in terms of the percentage of each class of indicators used on average by each index. In the CCI, the expert ratings by commercial risk agencies are the dominant class of indicator; in the CPI, the expert ratings by commercial risk agencies and the surveys of business executives practically overwhelm the other three classes of indicators (see the rows for "All Years" in Table 3). Second, there is considerable variation in terms of the reliance by each index on different classes of indicators over time. In the CCI, the relative number of expert ratings by commercial risk agencies declines markedly over time, as does the expert ratings by MDBs, while the relative number of expert ratings by NGOs and surveys of the mass public increases; in the CPI, the relative number of surveys of business executives declines very markedly over time, while the relative number of expert ratings by commercial risk agencies and the expert ratings by NGOs and MDBs increases (see Table 3). Third, the relative number of different classes of indicators varies considerably on average, that is, over all the years considered, across regions (see Table 4).

[8] Though the CPI uses, as a rule, at least three sources on each country (Lambsdorff 2008: 9), the CCI includes a country even if only one source is available (Kaufmann, Kraay and Mastruzzi 2008: 17). However, the number of countries included in the CCI with only one source is small.

**Table 3. Data Sources in the Control of Corruption Index (CCI) and the Corruption Perceptions Index (CPI) I: Classes of Indicators, by Year**

| Index | Year | Class of Indicators (% of all data sources) Expert Commercial | Expert NGO | Expert MDB | Business Survey | Mass Survey |
|---|---|---|---|---|---|---|
| CCI | All Years | 54.8 | 8.5 | 18.7 | 11.9 | 6.1 |
| | 1996 | 61.5 | 0.0 | 24.8 | 13.7 | 0.0 |
| | 1998 | 67.0 | 2.5 | 19.9 | 10.6 | 0.0 |
| | 2000 | 63.2 | 2.3 | 18.6 | 14.4 | 1.5 |
| | 2002 | 65.4 | 2.0 | 16.4 | 13.7 | 2.5 |
| | 2003 | 63.0 | 5.9 | 15.1 | 13.7 | 2.3 |
| | 2004 | 51.1 | 13.1 | 19.0 | 10.9 | 6.0 |
| | 2005 | 50.4 | 12.9 | 18.5 | 11.5 | 6.7 |
| | 2006 | 46.0 | 11.9 | 19.2 | 10.5 | 12.3 |
| | 2007 | 45.1 | 13.0 | 18.7 | 10.6 | 12.6 |
| CPI | All Years | 39.4 | 9.0 | 5.3 | 45.2 | 1.0 |
| | 2000 | 34.0 | 3.1 | 0.0 | 61.0 | 1.8 |
| | 2001 | 27.2 | 2.8 | 0.0 | 69.9 | 0.0 |
| | 2002 | 21.3 | 13.8 | 0.0 | 62.1 | 2.9 |
| | 2003 | 32.6 | 10.7 | 4.5 | 50.1 | 2.1 |
| | 2004 | 42.9 | 10.2 | 4.1 | 41.0 | 1.9 |
| | 2005 | 45.5 | 10.5 | 2.4 | 41.7 | 0.0 |
| | 2006 | 51.0 | 2.8 | 9.5 | 36.7 | 0.0 |
| | 2007 | 47.8 | 13.8 | 16.7 | 21.7 | 0.0 |

*Note:* The countries covered by both the CCI and the CPI are all the countries in the world covered by these indices. The N for the CCI for all years is 1890; the N for the CPI for all years is 1058. In 2006 and 2007, the CCI added one new indicator generated by experts within a government, which we include under the expert MDB type; this indicator accounts for 3.98% and 4.06% of the total number of sources in 2006 and 2007 respectively.

*Data Source:* Calculations based on a data set constructed from World Bank (2009a) and from data from Transparency International (2009).

**Table 4. Data Sources in the Control of Corruption Index (CCI) and the Corruption Perceptions Index (CPI) II: Classes of Indicators, by Region**

| Index | Region | Class of Indicators (% of all data sources) Expert Commercial | Expert NGO | Expert MDB | Business Survey | Mass Survey |
|---|---|---|---|---|---|---|
| CCI | North America | 65.6 | 3.3 | 2.2 | 20.0 | 8.9 |
| | Australasia | 70.4 | 2.0 | 1.3 | 23.7 | 2.6 |
| | Western Europe | 69.8 | 0.8 | 1.4 | 21.3 | 6.6 |
| | Eastern Europe & Baltics | 51.1 | 13.7 | 10.7 | 18.8 | 5.7 |
| | Post-Soviet | 47.9 | 19.4 | 17.0 | 12.0 | 3.6 |
| | East Asia | 57.4 | 4.8 | 9.7 | 23.9 | 4.2 |
| | S.E. Asia | 52.6 | 9.6 | 16.3 | 17.5 | 4.1 |
| | Pacific Islands | 35.0 | 1.8 | 62.5 | 0.0 | 0.7 |
| | South Asia | 49.3 | 11.1 | 27.2 | 8.3 | 4.2 |
| | MENA | 67.7 | 8.7 | 14.0 | 7.6 | 1.9 |
| | Sub-saharan Africa | 47.0 | 8.4 | 34.3 | 4.1 | 6.2 |
| | Latin America | 52.3 | 8.2 | 13.8 | 10.9 | 14.8 |
| | Caribbean | 62.9 | 3.3 | 25.5 | 5.0 | 3.3 |
| CPI | North America | 36.2 | 2.5 | 0.0 | 56.4 | 4.9 |
| | Australasia | 35.8 | 3.4 | 0.0 | 58.1 | 2.7 |
| | Western Europe | 28.1 | 3.2 | 0.0 | 65.4 | 3.2 |
| | Eastern Europe & Baltics | 29.4 | 18.1 | 1.9 | 50.4 | 0.1 |
| | Post-Soviet | 36.2 | 26.3 | 6.4 | 30.4 | 0.6 |
| | East Asia | 45.9 | 4.0 | 1.5 | 45.5 | 3.1 |
| | S.E. Asia | 45.9 | 5.3 | 3.6 | 44.2 | 1.0 |
| | Pacific Islands | 47.8 | 10.9 | 28.3 | 13.0 | 0.0 |
| | South Asia | 45.2 | 8.5 | 10.8 | 35.5 | 0.0 |
| | MENA | 54.7 | 8.6 | 3.6 | 33.1 | 0.0 |
| | Sub-saharan Africa | 43.6 | 9.5 | 17.2 | 29.7 | 0.0 |
| | Latin America | 36.5 | 7.6 | 2.7 | 53.2 | 0.0 |
| | Caribbean | 59.6 | 3.7 | 6.2 | 30.4 | 0.0 |

*Note:* For the CCI, the data are for 1996, 1998, 2000, and 2002-07; for the CPI, the data cover the 2000-07 years. The countries covered by both the CCI and the CPI are all the countries in the world covered by these indices. The N for the CCI for all years is 1890; the N for the CPI for all years is 1058. In 2006 and 2007, the CCI added one new indicator generated by experts within a government, which we include under the expert MDB type.

*Data Source:* Calculations based on a data set constructed from World Bank (2009a) and from data from Transparency International (2009).

The consequence of the lack of a balanced distribution of data sources across the different classes of indicators is that classes of indicators that are more frequently used have more weight. Thus, different classes of indicators are implicitly assigned different weights, both across countries as well as for a given country over time. And, since the indicators used by the CCI and the CPI are largely the same ones as those discussed in the previous section (i.e. indicators that diverge systematically in terms of their level of strictness and are not interchangeable), this differential weighting of classes of indicators due to the simple choice of indicators is potentially—unless addressed adequately when the indicators are aggregated—a significant threat to the validity of the CCI and the CPI.

**II. ii. The Choice of Aggregation Procedure**

The aggregation of indicators to form an index entails several choices, including the weights that are assigned to each indicator. Thus, the problem of differential weighting of classes of indicators that affects the CCI and the CPI in light of their choices of indicators could be remedied through a reweighting of indicators when the indicators are aggregated.[9] Indeed, the simplest way to counter the problem identified above with existing indicators of corruption is to weight the indicators in such a way that each class of indicator is given an equal weight in the index. Thus, it is critical to consider how the developers of the CCI and the CPI assign formal weights to indicators.[10]

---

[9] This is a standard practice in survey research. For example, if a certain group of a population was oversampled or undersampled, a post-survey weighting is conducted to solve the problem. In the case of the data sources used in corruption indices, such a solution is hampered by the lack of available indicators. Indeed, a balanced index would only be possible for those countries where all main classes of indicators are available and hence involve a loss of empirical scope. But the matter of empirical scope of an index should be kept separate from an assessment of the validity of an index.

[10] Beyond the decision about what weight to assign each indicator that has been selected for an index, the aggregation of the indicators involves two other questions: (1) what, if any, rescaling of the original indicator scales, is called for? and (2) what is the relationship among the indicator values being aggregated? With regard to the second question, both the CCI and the CPI rely on an additive aggregation rule. This is a standard default option, which largely makes sense in this context. In particular, since the developers of the CCI and the CPI posit that each indicator they aggregate is simply a different measure of the same concept, as opposed to say a part of a whole, the use of an additive aggregation rule is a relatively uncontentious matter. With regard to the first question, the CCI and the CPI proceed differently. The CCI relies on a transformation of all the indicator values into normalized scores using a 0-1 scale, a fairly standard if not entirely justified procedure. In contrast, the CPI is based on the more modest rule of relying only on the information about ordinal differences captured by the indicator values (Lambsdorff 2008: 6). However, from the perspective of our analysis, this difference has little impact. Thus, we focus solely on the question of weighting of indicators. For a detailed discussion of aggregation procedures, see Lambsdorff (2006b: 88-97; 2008: 6-8) on the CPI, and Kaufmann, Kraay and Mastruzzi (2004: 258-61, 2008: 13-16, 97-102) on the CCI.

The CPI uses a simple average of all the indicators used for each country-year and thus formally assigns equal weight to each of the available indicators (Lambsdorff 2006b: 97). However, because the class of indicators used in the CPI vary considerably over time and across countries (see Table 3 and 4), this rule actually implies that indicators of different classes will, in effect, be weighted differently both across countries and over time. In other words, the choice of weighting of indicators used in the CPI ignores the problem that different classes of indicators vary systematically in their assessments—offering stricter or more lenient evaluations of the same countries. As a result, CPI country scores will inevitably be due, in part, to a nonrandom factor, whether the distribution of classes of indicators for a particular country-year tilts more toward stricter or more lenient indicators. In short, the default choice of averaging the values of indicators used in the CPI does not resolve the problem associated with using data on indicators of corruption that are characterized by systematic measurement error.

The weighting of indicators in the CCI is more complex, in that it involves a conscious choice of assigning variable weights to indicators. Specifically, the CCI relies on the unobserved component model to assign variable weights to all the available indicators used m the CCI (Kaufmann, Kraay and Mastruzzi 2008: 97-102).[11] Thus, the CCI uses a data-driven rather than a theory-driven procedure to determine the weight of indicators. The indicator weights are determined by an empirical analysis of the data on the selected indicators rather than being explicitly assigned, on theoretical grounds. But the unobserved component model inescapably relies on certain assumptions, that must be justified and that, in the case of the CCI, can be challenged.

First, the unobserved component model assumes that the indicators that are analyzed offer independent measures and, more specifically, that the error terms of the indicators are not correlated. Yet there is at the very least anecdotal evidence that suggests that there is contamination across indicators because the evaluators used to generate data on certain indicators are influenced by the data on other indicators (Arndt and Oman 2006: 65-66, Galtung 2006: 116, Knack 2007: 266-70). Thus, though arguments in favor of this assumption can certainly be made (Kaufmann, Kraay and Mastruzzi 2006: 19-31, 42-44; 2007c: 16-20; see also Lambsdorff 2007: 245), it is unlikely to hold in the context of the measurement of corruption.

[11] After this weighting of indicators has been carried out, the values of all the available indicators are averaged.

Second, and more fundamentally, the unobserved component model assumes that more highly correlated indicators should be assigned a greater weight because such indicators offer truer values. But this assumption can be questioned both on methodological and empirical grounds. Methodologically, the assumption is flawed because it mistakenly uses a test of the reliability of data as a test of validity or, in other words, consensus for truth. Interpreting correlations between measures as evidence of validity of these measures confuses the concept of validity and reliability. Validity concerns the relationship between a concept and its measure and thus must inevitably call for references back to the abstract concept that is purportedly being measured, whereas reliability focuses on the level of agreement among multiple measures but does not distinguish among reliably correct measures and reliably incorrect measures.

In turn, empirical tests also caution against the use of the unobserved component model. As shown in the analysis of indicators provided above, different classes of indicators exhibit systematic measurement error. Thus, one of the key concerns in constructing a corruption index should be, as emphasized, the need for a balanced distribution of data sources across the different classes of indicators. Yet the effect of the unobserved component model is precisely the opposite. The classes of indicators that are more frequently selected for inclusion in the CCI are, due to their sheer number, more heavily weighted. But the classes of indicators that are more frequently selected for inclusion in the CCI are also more highly correlated,[12] and hence on average are assigned a (non-trivial) heavier weight—the weights assigned to indicators through the unobserved component model range, for the same year, from 0.2239 to 0.0001 (World Bank 2009b). Thus, the methodological misgivings regarding the unobserved component model are, in this context, reinforced by an empirical analysis of the indicators.

In short, the CCI's more complex approach to the weighting of indicators not only does nothing to counteract the problem that the CCI is based, first of all, on indicators that contain systematic measurement error. Furthermore, the CCI's formal weighting scheme actually

---

[12] As Table 3 shows, the most frequently used classes of indicators in the CCI are, in descending order, the expert ratings by commercial risk assessment agencies, the expert ratings by MDBs, the surveys of business executives, the expert ratings by NGOs, and the surveys of the mass public. In turn, the average level of correlation between any single indicator of a certain class of source indicator and all other indicators, a gauge of the level of agreement among indicators, is as follows: expert ratings by commercial risk assessment agencies = 0.552, expert ratings by MDBs = 0.425, surveys of business executives = 0.526, expert ratings by NGOs = 0.373, and surveys of the mass public = 0.323. Thus, to a large extent, the most frequently used classes of sources indicators are also more highly correlated. Correlations were calculated for all 25 indicators used in the CCI (World Bank 2009a). See Appendix 1 for the classification of indicators into different classes of sources indicators (for the purpose of this test, the data by the Institutional Profiles Database was combined with the expert ratings by MDBs indicators).

compounds the imbalance in the distribution of data sources across the different classes of indicators due to the selection of indicators. Hence, the CCI's weighting scheme is even more problematic than the simpler weighting scheme adopted by the CPI.

### II. iii. The Impact of Alternative Weightings of Classes of Indicators

The effect of the choice of indicators and the choice of aggregation procedure on the validity of the CCI and the CPI cannot be quantified. Such an estimate would require knowledge of the true level of corruption around the world, which is precisely the aim of work on measurement. But we can isolate variation in the values of the CCI and the CPI that is due solely to a methodological choice, the variable weights they assign to different classes of sources both through the choice of indicators and the choice of aggregation procedure. And such an assessment of the impact of alternative weights of classes of indicators is instructive.

To this end, we devised the following test. First, we used the data on country-year for which indicators of all five classes of sources are available to calculate an average value for indicators of each class of source.[13] Second, we used information about what indicators were used by the CCI and the CPI for each country-year,[14] and we input in each data point that used an indicator of a certain class the corresponding average value previously calculated. Third, using the new data for each indicator, we used the CCI's and the CPI's formal weighting scheme to calculate the value of these indices. In this way, we arrived at values for two indices that, by holding the values of each class of indicator constant, only differ in terms of the weight assigned to each indicator (either informally through the variable number of indicators of different classes used in an index or formally through the weights assigned to each indicator). In other words, for the sake of this test, we posit that, according to each class of indicator, all countries have the same level of corruption (only different classes of indicators differ in their assessment), so that any variation in the value of an index can be directly and unequivocally attributed to variable weight assigned to each class of indicators.

---

[13] The average value of each class of indicator was calculated by first computing an average across indicators of a certain class for each country-year and then averaging the values for each indicator class across all country-years. These values are as follows: expert ratings by commercial risk assessment agencies = 0.35, surveys of business executives = 0.44, expert ratings by MDBs = 0.48, expert ratings by NGOs = 0.50, and surveys of the mass public = 0.50. These values were calculated using all 25 indicators used in the CCI (World Bank 2009a) and an N of 250. For our classification of indicators into different classes of indicators, see Appendix 1 (the one exception is that for the purpose of this test, the data from the Institutional Profiles Database was combined with the expert ratings by MDBs).

[14] For the CCI, the years covered are 1996, 1998, 2000, and 2002-07; for the CPI, 2000-07. The number of countries varies, reaching a maximum of 210 per year for the CCI and 180 per year for the CPI.

The results of this test are telling. A comparison of the means of the different indices shows that differences in weighting schemes are not irrelevant. The mean values of indices using the CCI's and the CPI's weighting scheme, as would be expected from the relatively larger weight they assign to evaluators that are stricter on average (see Tables 1 and 3), are significantly lower than the mean value of an index that assigns an equal weight to the five classes of indicators identified in this paper (see Table 5).[15] Of course, the lower means could be the more accurate measure. But, in the absence of firm evidence that certain classes of indicators are more valid than others, there is no theoretical reason to give certain classes of indicators more weight than others.[16] Thus, this result shows that the values of the CPI, and even more so those of the CCI, diverge from the most reasonable weighting of indicators we can currently make.[17]

**Table 5. Alternative Weightings of Classes of Indicators and Methods-induced Bias**

| Index Weighting Scheme * | Mean *** | Standard Deviation | Min (Deviation from Mean) | Max (Deviation from Mean) |
|---|---|---|---|---|
| Equal Weighting ** | 0.454 | 0.000 | 0.000 | 0.000 |
| CPI Weighting | 0.410 | 0.021 | -0.060 | 0.080 |
| CCI Weighting | 0.396 | 0.033 | -0.046 | 0.104 |

*Note:* (*) So as to limit variation to the weighting scheme, the same values for each class of indicator were used in calculating all indices; these values were calculated using all 25 indicators used in the CCI; N = 250.

(**) Equal weight was assigned to indicators of the following five classes: expert ratings by commercial risk assessment agencies, expert ratings by MDBs, expert ratings by NGO, surveys of business executives, and surveys of the mass public.

(***) All scales are 0.000 to 1.000 scales, with higher values being indicative of a lower level of corruption.

The N for the equal weighting row is 1878, for the CPI weighting row 1057, and for the CCI weighting row 1787.

*Source:* Authors' calculations based a data set constructed from World Bank (2009a) and information about sources from Transparency International (2009).

[15] The mean value of an index is simply the average of the values of the five classes of indicators.
[16] This point has been made by Knack (2007: 268-69).
[17] The deleterious impact of the CCI's reliance on the unobserved component model in particular is obvious when an index using the full CCI weighting is compared to an index that averages the values of the indicators used by the CCI and hence drops the extra weighting introduced via the unobserved component model. The index with a simple weighting has an average of 0.405, standard deviation of 0.032, a minimum deviation from the mean of -0.055 and a maximum deviation from the mean of 0.095. In other words, by aggregating indicators using the unobserved component model the CCI is more unbalanced than if indicators were aggregated using a simple averaging formula.

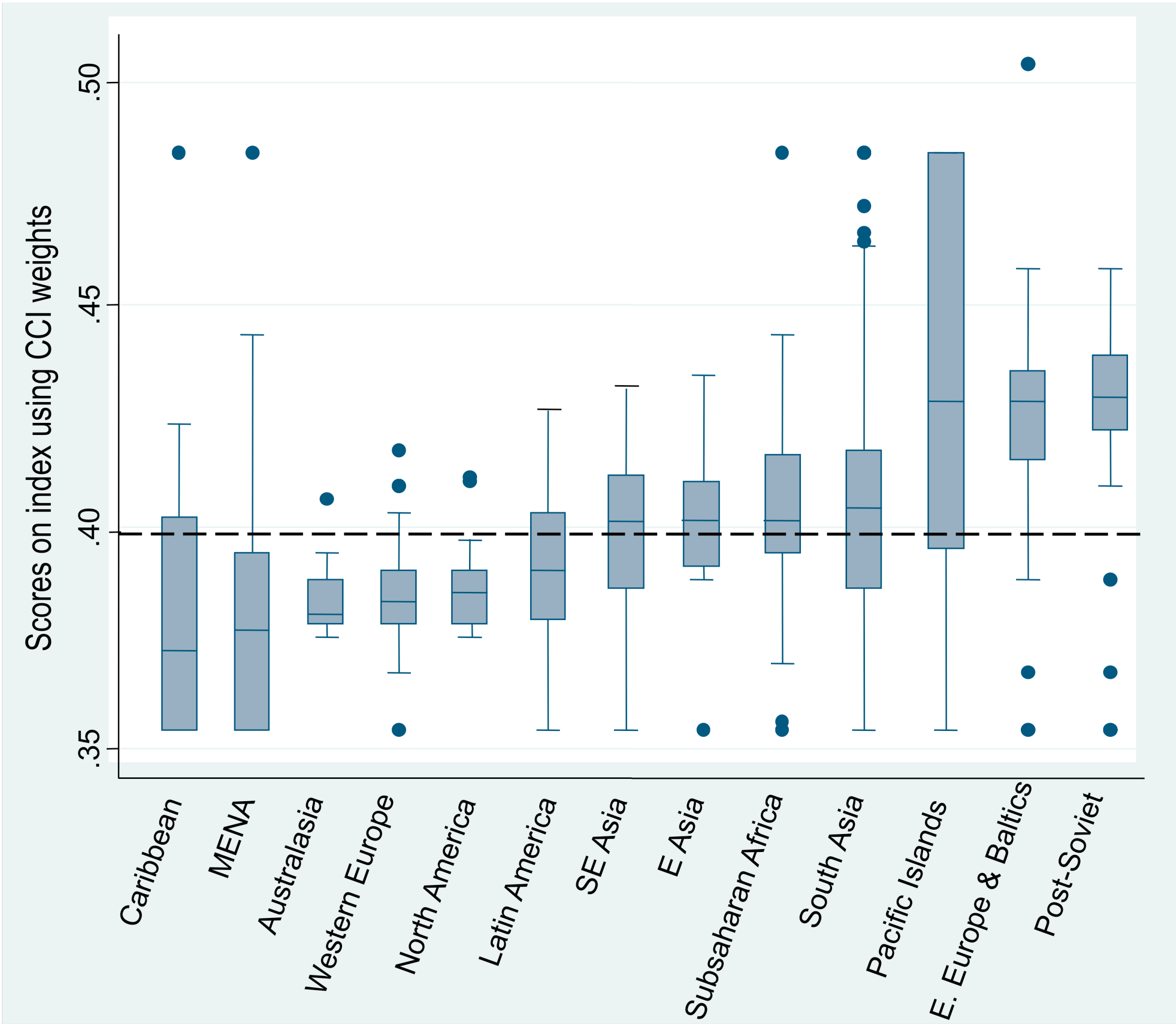
The information on the spread of the values of the CCI and the CPI around their means is even more striking, in that it relies on a purely methodological baseline. Given the design of our test, and the fact that the indicator values are *exactly* the same for each country, if the weights assigned to the indicators were consistently applied to every country, all countries would receive the same scores. And this is indeed what happens with the index based on the equal weighting of the five classes of indicators: the standard deviation is 0, and the minimum and maximum values do not deviate from the mean value. But the indices calculated using the CCI's and the CPI's choice of indicators and formal weighting scheme diverge considerably from their mean values. In other words, a significant part of the variation in the CCI and the CPI is solely due to the lack of a consistent weighing of classes of indicators and has no relation to variation in levels of corruption in the real world.

These methodological errors are not innocent. For example, though the mean values of the CCI and CPI in this test should remain constant over time, the index using the CCI's weights points to a *reduction* in the level of corruption, from 0.382 in 1998 to 0.415 in 2006, while the index using the CPI's weights shows an *increase* in the level of corruption, from 0.432 in 2002 to 0.402 in 2006.[18] In turn, the lack of a consistent weighing scheme on the part of the CCI and the CPI leads to systematic fictitious differences across the main world regions, the magnitude of these differences being larger in the case of the index using the CCI's weights (see Figures 1 and 2). In short, the construction of the CCI and the CPI lacks a solid theoretical justification—it seems to place the desire to cover the largest number of countries possible above key methodological considerations—and induces bias in the CCI and CPI data that seriously undermines their usefulness in comparing levels of corruption across countries and over time.[19]

[18] Both difference of means are significant at a 0.0000 level. The N for the index with CCI weights is 182 per year; for the index with CPI weights the N is 80 per year.

[19] It may seem surprising that a test reported by the developers of the CCI shows that the values of the CCI are robust to changes in the weighting scheme and, specifically, that an index using the weighting scheme chosen for the CCI is very highly correlated to one that gives equal weight to different classes of indicators. Specifically, Kaufmann, Kraay and Mastruzzi (2007c: 21-22) report correlations for all six World Governance Indicators, of which the Control of Corruption Index is one, and state that the correlation between an index using their weighting scheme and one that weights each class of indicator equally is approximately 0.95. But a simple analysis of correlations does not constitute an adequate test of the impact of these different weighting schemes on the data on corruption provided by the World Bank or Transparency International for a simple reason. An index that gives equal weight to different classes of indicators must, perforce, be limited to those countries for which all classes of indicators are available. And precisely because all classes of indicators are available for these countries, the weight of each of class of indicators will be more balanced in these countries than in the countries that are not used in the correlational analysis. In short, such a test has to rely on a subset of cases that is not representative of the entire

**Figure 1. Regional Variation in Methods-induced Bias I:**
**The World Bank's Control of Corruption Index (CCI)**




*Note:* This graphic shows the methods-induced deviations in corruption scores that result from differences across countries in the weighting of classes of indicators. To illustrate these methods-induced deviations in corruption scores, we begin by assuming that the corruption experience in all countries is homogenous and assigning each country the same score for any indicator of the same class of indicator. These scores are the actual mean value of indicator of a certain class for all countries that have indicators of all five classes of sources (see footnote 13 for more details). We then calculate the CCI index score for each country, using the list of indicator used in the CCI and the indicator weights provided by the World Bank.

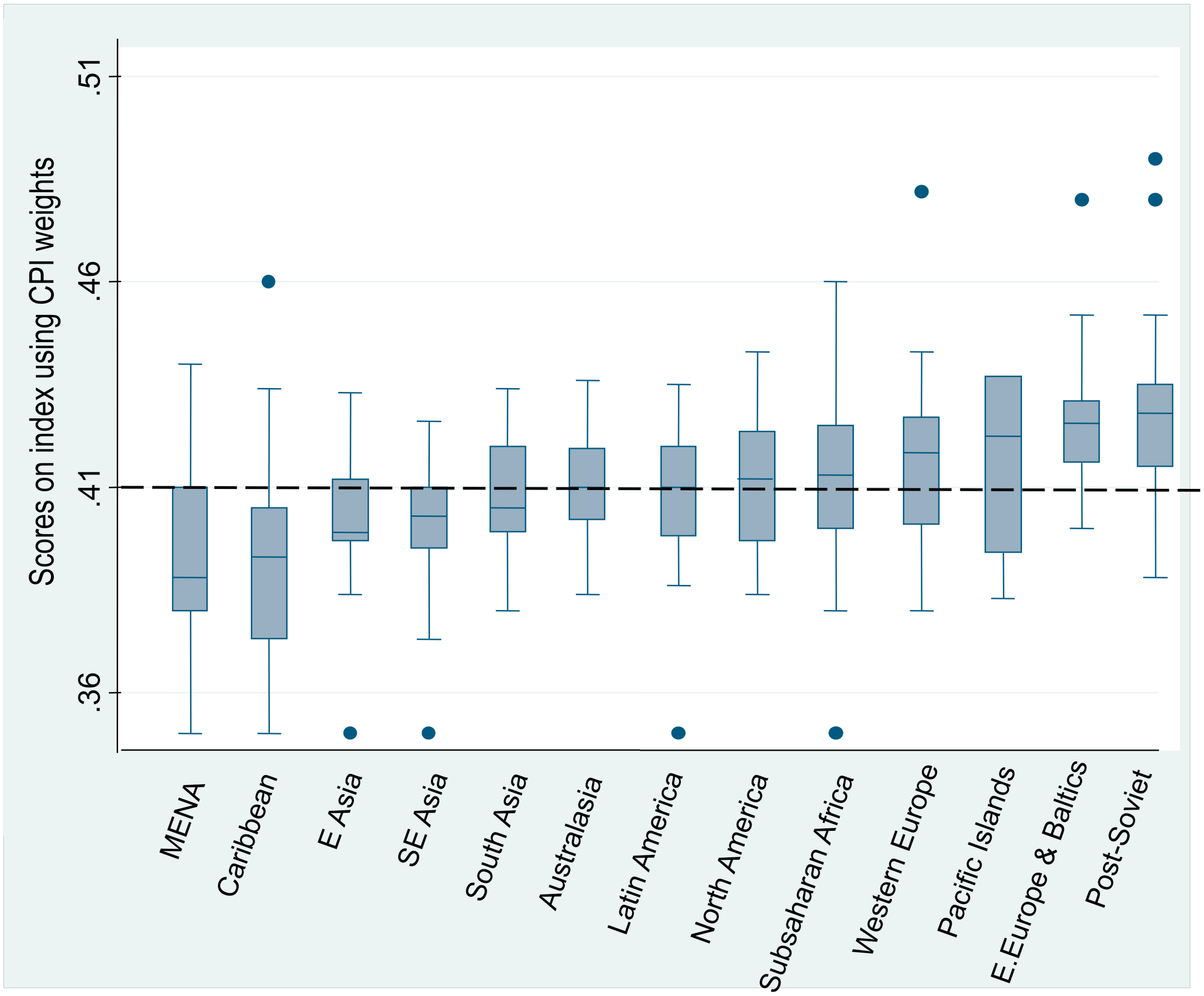
The heavy dashed line indicates the mean score across all countries. If there were no methods-induced deviations in scores, there would be no variation across countries, and all scores would align on the heavy dashed line, which represents the global mean of 0.396. Deviations from this mean show how much a country's score is shifted upward or downward due to the variable weighting of classes of indicators. Countries with deviations that fall above the mean line benefit from methods-induced improvements to their corruption scores; countries that fall below the mean line, suffer a methods-induced decrease in their corruption scores. These box plots show that the size of the methods-induced deviations in corruption scores differs by region. For a list of the countries included in each region, see Appendix 2.

*Source:* Authors' calculations based a data set constructed from World Bank (2009a).

---

population, and hence does not offer an adequate estimate of the impact of different weighting schemes on the CCI or CPI data.

**Figure 2. Regional Variation in Methods-induced Bias II:**
**Transparency International's Corruption Perception Index (CPI)**




*Note:* See the explanatory note in Figure 1. The heavy dashed line indicates the mean score across all countries. If there were no methods-induced deviations in scores, there would be no variation across countries, and all scores would align on the heavy dashed line, which represents the global mean of 0.410.

*Source:* Authors' calculations based a data set constructed from World Bank (2009a) and information about sources from Transparency International (2009).

## III. Implications of the State of Corruption Measures for Future Research

Efforts to measure corruption, and especially to develop cross-national data sets of broad scope, are unquestionably welcome and laudable. Corruption is a key problem that directly undermines democracy and good government. And part of the challenge of understanding the nature of corruption, as well as its causes and consequences, is the development of measures of corruption. Thus, it is always worth remembering that having some data, even if of poor quality, is a less grave problem than having no data at all. Yet the current availability of multiple indicators and a few indices on corruption has given rise to a new problem, that different measurement methodologies produce different measures of corruption and, as shown, that a substantial amount of the variation in reported levels of corruption is an artifact of the methodologies used to generate the data as opposed to variation in actual levels of corruption. This is a sobering conclusion, which has several implications both for data users and producers.

*Implications for Data Users*. For users of data, the most generic advice that can be offered is that, in light of the finding that the choice of data set has a considerable impact on the results of quantitative, cross-national studies on corruption, researchers need to pay more attention to data matters. As shown, different datasets support divergent descriptions of the world. And, unsurprisingly, different data sets support different findings about the causes and consequences of corruption. Indeed, our review of 76 articles published in 19 highly regarded economics journals between 1995 and 2009[20] reveals that when hypotheses were tested with multiple data sets on corruption, even though these data sets were not always independent ones (i.e. they are indices that use the same data sources), the coefficient changed sign in 28% of the tests and the magnitude of the key coefficient varied considerably in many of the other cases. We also found that the strength of statistical conclusions were very sensitive to choice of data set: the significance level of the coefficients changed (crossing either the 5% or the 10% threshold) in 60% of the tests.[21] Thus, researchers need to be aware of the importance of their choice of corruption measure and familiarize themselves with the range of data sets available. Indeed, before undertaking an empirical analysis every researcher would do well to ask him or herself, "Do you know your data?"

More specifically, this paper has implications for researchers seeking to navigate their way through the maze of available corruption measures. Our analysis does not provide a sufficiently clear basis for distinguishing the relative validity of the multiple proposed measures

[20] For the full discussion of the data set, see Angela Hawken and Gerardo L. Munck, "Data Set on Corruption Research in Economics Journals: Case Selection, Coding and Variable Description," May 28, 2009 (available upon request).

[21] In 47% of the cases the results were above and below the 5% threshold of significance, and in 37% of the cases the results were above and below the 10% threshold of significance. Since some of the articles include tests of more than one research question, these percentages are based on the total number of research questions that are tested. Authors' calculations based on the *Hawken and Munck Data Set on Corruption Research in Economics Journals* (2009).

of corruption and for arguing unequivocally that researchers should opt for one measure over others. To produce a ranking of measures, further information is needed, such as the results of intercoder reliability tests for the indicators that use expert rating; the criteria of selection of respondents, response rates and levels of disagreement among survey respondents; and all the underlying data used in indices. Data producers do not make this information publicly available and it is not even clear whether all of this information has even been collected. Nonetheless, some conclusions relevant to the use of measures of corruption deserve to be highlighted.

First, with regard to indicators of corruption, that different evaluators use different standards in evaluating levels of corruption and that the difference in standards across evaluators does not hold consistently across countries is an indication that the measurement of corruption remains an imprecise enterprise. Users of the indicators of corruption should be very caution about drawing any strong conclusions from empirical analysis that relies on these measures. It is crucial that researchers recognize the need for sensitivity analyses geared to testing the robustness of their findings to different measures and make such tests a standard element in their analyses, much as is the case with regard to tests that address alternative specifications of a causal model. Yet conducting sensitivity analyses that address the impact of the choice of data set is not an established practice. Our review of articles recently published in economics journals shows that only 39% of authors who used corruption data sets performed a sensitivity analysis in which a hypothesis was tested using more than one data set. In turn, a sensitivity analysis that both held constant the specification of the model and relied on partly or fully independently generated data sets was offered in only 25% of the reviewed articles.[22] Thus, along with emphasizing the need for sensitivity analyses, it is key to stress that the data sets included in such an analysis should be independent data sources and, preferably, should draw on data sources on at least two of the five classes of indicators identified in this paper, that is, expert ratings by commercial risk assessment agencies, by NGOs and by MDBs, and surveys of business executives and the mass public.

Second, though combining indicators to generate an index can reduce the measurement error associated with indicators, the problem with indicators on corruption—that measures of corruption vary according to who the evaluator is—is imported into the Transparency

[22] Authors' calculations based on the *Hawken and Munck Data Set on Corruption Research in Economics Journals* (2009).

International's CPI and the World Bank's CCI. Thus, these indices do not solve the problem of corruption indicators and are not obviously preferable to them.[23] The authors of the CCI recognize the inherent uncertainty in their reported index values and publish error estimates along with their index (Kaufmann, Kraay and Mastruzzi 2008: 1-2, 17). However, such estimates depend on questionable assumptions—that the indicators that are combined are independent and that more highly correlated indicators more closely approximate the true value—and ignore other sources of measurement error—such as whether the indicators actually measure corruption as opposed to some extraneous concept and whether the choice of indicators and formal weighting of indicators introduces any bias in the way variation is tracked over time and across countries. Both the CPI and the CCI give a false sense of precision which overstates their usefulness, as some critics have pointed out, for the purposes of either time-series analysis (Arndt and Oman 2006: 61, 67-69, Knack 2007: 264-65) or cross-sectional analysis (Arndt and Oman 2006: 60, Iqbal and Shah 2008: 35-39).

It is hard to say whether the CPI or the CCI is less problematic. Overall, the developer of the CPI demonstrates greater awareness about issues of validity than the developers of the CCI.[24] In particular, the CPI's indicators were more clearly selected to measure corruption and avoid the surreptitious inclusion of extraneous factors, making the CPI conceptually sounder than the CCI (compare Lambsdorff 2007: 238-44 to Kaufmann, Kraay and Mastruzzi 2007d: 75). Moreover, the CPI relies on a less problematic weighting scheme, and its developer is more transparent about the limitations of this index.[25] But both the CPI and the CCI suffer from methodological weaknesses and have not been appropriately validated by their developers. And, since Transparency International does not make their underlying data public, a major drawback of the CPI, independent researchers are unable to conduct tests on the CPI and carry out a full

[23] This is a key point because the quantitative literature relies very heavily on the Transparency International's CPI, and the World Bank's CCI; they are used in 45% and 24% of all articles, respectively. The other popular data set is the commercial Political Risk Services' International Country Risk Guide (ICRG), used in 58% of all articles. Authors' calculations based on the *Hawken and Munck Data Set on Corruption Research in Economics Journals* (2009).

[24] Indeed, even when critics of the developers of the CCI have pointed out their failure to clearly distinguish these two concepts (Kurtz and Schrank 2007b: 564), their response does not even acknowledge the point (Kaufmann, Kraay and Mastruzzi 2007b). Nonetheless, the developer of the CPI also confuses the concepts of validity and reliability quite often. An illustrative example is when Lambsdorff (2006b: 87), after considering the correlation among various indicators, states that "The validity of the sources is mutually confirmed."

[25] For example, the developer of the CPI, Lambsdorff (2006b: 83-84) admits that the CPI is not well suited for over-time comparisons. In contrast, the developers of the CCI reject out of hand the value of any criticism of the CCI (Kaufmann, Kraay and Mastruzzi 2007c, Kaufmann and Kraay 2008: 21-22).

comparison between the CPI and the CCI. A sensitivity analysis using both the CPI and the CCI suffers the problem that both indices use many of the same indicators and hence are not really independent measures of corruption (indeed, the shortcomings of sensitivity analyses that use one of these indices and one of the indicators included in the selected index should also be recognized). Thus, researchers must weight the convenience of ready accessibility to data on a large number of countries—undoubtedly the key selling point of the CPI and the CCI—against the important limitations of these measures.

*Implications for Data Producers*. It is important to recognize that efforts to generate new measures and to improve the quality of data on corruption are an integral part of corruption-oriented research. One relatively easy way to improve on the available measures of corruption is to develop a superior index from the available data on indicators. A clear advance would be made by generating an index that better addresses the two key methodological choices raised in this paper—the choice of indicators and aggregation procedure—and that is based on choices guided both by theory and empirical tests. Such an index would perforce be narrower in empirical scope than the CPI and the CCI if it required balance among the classes of indicators used for each country. Yet it may well be preferable to test theories about the causes and consequences of corruption with a smaller N than is provided by the CPI and the CCI but with greater certainty about the validity of the data.

A more burdensome task concerns the generation of data on new indicators. In this regard, the measurement literature has emphasized the need to develop "objective" or hard measures as opposed to the "subjective" or soft measures of corruption based on evaluators. The distinction between objective and subjective measures of corruption is relevant, in that research has shown that these two classes of measures produce divergent assessments (Mocan 2008, Donchev and Ujhelyi 2009, Olken 2009). Moreover, one clear advantage of objective measures is that they avoid a potential critique of subjective measures, that evaluators assess the level of corruption in a country at least in part based on suspected causes and consequences of corruption, such as poverty and economic growth, instead of corruption itself.[26] Thus, the steps

[26] This problem has been labeled the halo effect in the literature on the measurement of corruption (Glaeser et al. 2004: 276, Kurtz and Schrank 2007a: 543, 2007b: 567).

taken to develop "objective," experience-based measures of corruption by a number of researchers is a welcome development.[27]

It is important to recognize that the development of "objective" corruption measures for the purpose of cross-national comparison is a daunting task. On the other hand, it is critical not to underestimate the potential of subjective measures of corruption. The contrast between objective and subjective measures is frequently exaggerated, and subjective measures are essentially equated with perceptions of corruption. So-called subjective measures may tap solely into perception, but they can also be carefully grounded in observables and even meet the basic criteria of replicability. Thus, future research on corruption will best be served by various classes of measures of corruption and by judiciously pooling the information from these different measures.[28]

The challenges discussed in this paper are not limited to research on corruption. To be sure, **c**orruption is a notoriously hard concept to measure. But many key concepts in political science and economics—such as democracy, human rights, justice, rule of law, state capacity, and development—are similarly hard to measure. And much of the data used in research on these topics are based, as in the case of corruption, on evaluations by a range of respondents, be they experts or common citizens, and thus are face the same problems found in measures of corruption. Indeed, measures of these concepts will, for the foreseeable future, be affected by the problem that evaluators rely on different standards of assessment and furthermore do not apply standards consistently across cases. [29] Thus, the question we address in this paper—does the evaluator make a difference?—has broad implications for the knowledge claims made in the social sciences.

---

[27] See Di Tella and Schargrodsky (2003), Reinikka and Svensson (2004), Golden and Picci (2005), Hsieh and Moretti (2006), and Olken (2007).

[28] To improve the quality of subjective measures, greater attention has to be given to the way in which questionnaires are written and administered. Questions should focus more on specifics and actual observations. Moreover, it is important not to ask people about things they really do not know about (for example, asking common citizens about grand corruption seems like an invitation to speculation) and to do more to tap into the knowledge of people who do have useful information (for example, by posing questions to people who work within government ministries). For an example of data gathered through a survey of ministries administered in Georgia, see Hawken and Kulick (2006). For an instrument, designed for Afghanistan, that uses a survey of ministries, and combines it with other sources of information, see Hawken and Munck (2008). Finally, for an instrument to generate cross-national measures using different sources of information, see Hawken and Munck (2009).

[29] For some evidence of a "coder factor" in measures of democracy, see Munck (2009: 77-79).

**Appendix 1: Sources of Data on Indicators**

| Class of Indicator | Data Source | Name of Measure * |
|---|---|---|
| Expert Rating/ Commercial Risk Assessment Agency | Business Environment Risk Intelligence (BERI) | Business Risk Service |
| | Business Environment Risk Intelligence (BERI) | Qualitative Risk Measure in Foreign Lending |
| | Economist Intelligence Unit | Country Risk Service, Country Forecasts |
| | Global Insight | Business Conditions and Risk Indicators |
| | Global Insight | Global Risk Service |
| | Merchant International Group | Grey Area Dynamics |
| | Political Risk Services | International Country Risk Guide |
| Expert Rating/ Non-governmental Organization (NGO) | Bertelsmann Foundation | Bertelsmann Transformation Index |
| | Freedom House | "Anticorruption and Transparency" measure in *Countries at the Crossroads* |
| | Freedom House | "Corruption" measure in *Nations in Transition* |
| | Global Integrity | Global Integrity Index ** |
| Expert Rating/ Multilateral Development Bank (MDB) | African Development Bank | Country Policy & Institutional Assessment (CPIA) |
| | Asian Development Bank | Country Policy & Institutional Assessment (CPIA) |
| | International Fund for Agricultural Development | Rural Sector Performance Assessments |
| | World Bank | Country Policy & Institutional Assessment (CPIA) |
| Expert Rating/ Governments | Institutional Profiles Database | Country Security Risk Ratings |
| Survey of Business Executives | Institute for Management Development (IMD) | "Executive Opinion Survey" in the *World Competitiveness Yearbook* |
| | Political Economic Risk Consultancy (PERC) | Corruption in Asia Survey |
| | World Bank | Business Enterprise Environment Survey (BEES) |
| | World Economic Forum (WEF) | "Global Competitiveness Survey" in the *Global Competitiveness Report* ** |
| Survey of the Mass Public | Afrobarometer | Afrobarometer Survey ** |
| | The Gallup Organization | Gallup World Poll |
| | Latinobarometro | Latinobarometro Survey |
| | Transparency International | Global Corruption Barometer ** |
| | Vanderbilt University | Americas Barometer Survey |

*Note:* (*) For information on the actual indicators, see Kaufmann, Kraay and Mastruzzi (2008: 78). (**) These "indicators" are actually the average of two or more indicators.

The data cover 210 countries over the 1996, 1998, 2000, and 2002-07 years. All of the data were accessed from the World Bank's Worldwide Governance Indicators website, where the data are made available as normalized scores using a 0-1 scale, with a 1 meaning a positive situation. There are gaps in the data for some countries for the CPIA scores of the three multilateral development banks.

*Source:* World Bank (2009a).

**Appendix 2: List of Countries and Territories by Region**

| Region | Countries and Territories |
|---|---|
| **Australasia** | Australia, New Zealand |
| **Caribbean** | Anguilla, Antigua and Barbuda, Aruba, Bahamas, Barbados, Belize, Bermuda, Cayman Islands, Dominica, French Guiana, Grenada, Guyana, Haiti, Jamaica, Martinique, Netherlands Antilles, Puerto Rica, St. Kitts and Nevis, St. Lucia, St. Vincent and the Grenadines, Suriname, Trinidad and Tobago, Virgin Islands (U.S.) |
| **E Asia** | China, Hong Kong, Japan, North Korea, South Korea, Macao, Mongolia, Taiwan |
| **Eastern Europe and Baltics** | Albania, Bosnia-Herzegovina, Bulgaria, Croatia, Czech Republic, Estonia, Hungary, Kosovo, Latvia, Lithuania, Macedonia, Poland, Romania, Slovakia, Slovenia |
| **Latin America** | Argentina, Bolivia, Brazil, Chile, Colombia, Costa Rica, Cuba, Dominican Republic, Ecuador, El Salvador, Guatemala, Mexico, Nicaragua, Panama, Paraguay, Peru, Uruguay, Venezuela |
| **MENA (Middle East and North Africa)** | Algeria, Bahrain, Chad, Egypt Iran, Iraq, Israel, Jordan, Kuwait, Lebanon, Libya, Morocco, Oman, Qatar, Saudi Arabia, Syria, Tunisia, Turkey, United Arab Emirates, West Bank/Gaza, Yemen |
| **North America** | Canada, United States |
| **Pacific Islands** | American Samoa, Cook Islands, Fiji, Guam, Kiribati, Marshall Islands, Micronesia, Nauru, New Caledonia, Niue, Palau, Papua New Guinea, Reunion, Samoa, Solomon Islands, Tonga, Tuvalu, Vanuatu |
| **Post-Soviet states (except the Baltics)** | Armenia, Azerbaijan, Belarus, Georgia, Kazakhstan, Kyrgyzstan, Moldova, Russia, Tajikistan, Turkmenistan, Ukraine, Uzbekistan |
| **South Asia** | Afghanistan, Bangladesh, Bhutan, India, Maldives, Nepal, Pakistan, Sri Lanka |
| **SE Asia** | Brunei, Cambodia, Indonesia, Laos, Malaysia, Myanmar, Philippines, Singapore, Thailand, Timor-Leste, Vietnam |
| **Subsaharan Africa** | Angola, Benin, Botswana, Burkina Faso, Burundi, Cameroon, Cape Verde, Central African Republic, Comoros, Congo, Congo Democratic Republic, Cote D'Ivoire, Djibuti, Equatorial Guinea, Eritrea, Ethiopia, Gabon, Gambia, Ghana, Guinea, Guinea-Bissau, Kenya, Lesotho, Liberia, Madagascar, Malawi, Mali, Mauritania, Mauritius, Mozambique, Namibia, Niger, Nigeria, Rwanda, Sao Tome and Principe, Senegal, Seychelles, Sierra Leone, Somalia, South Africa, Sudan, Swaziland, Tanzania, Togo, Uganda, Zambia, Zimbabwe |
| **Western Europe** | Austria, Belgium, Cyprus, Denmark, Finland, France, Germany, Greece, Iceland, Ireland, Italy, Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, United Kingdom |

*Note:* Other countries or territories included in the analysis but excluded from the categorization of regions are: Andorra, Liechtenstein, Luxemburg, Malta, Monaco, and San Marino.

## References


ADB (Asian Development Bank). 2006. “2005 Annual Report on ADB’s Country Performance Assessment Exercise.” Available at: http://www.adb.org/Documents/Reports/Country-Performance-Assessment-Exercise/default.asp

AfDB (African Development Bank). 2006. “The African Development Bank’s 2005 Country Performance Assessment. A Summary of the AfDB 2005 CPIA Ratings Scores.” Available at: http://www.afdb.org/en/documents/project-operations/country-performance-assessment-cpa/

Alesina, Alberto, and Beatrice Weder. 2002. “Do Corrupt Governments Receive Less Foreign Aid?” *American Economic Review* Vol. 92, Nº 4: 1126-37.

Arndt, Christiane, and Charles Oman. 2006. *Uses and Abuses of Governance Indicators* (Paris: OECD Development Centre Studies).

Di Tella, Rafael and Ernesto Schargrodsky. 2003. “The Role of Wages and Auditing during a Crackdown on Corruption in the City of Buenos Aires,” *Journal of Law and Economics* Vol. 46, Nº 1: 269–92.

Donchev, Dilyan and Gergely Ujhelyi. 2009. “What Do Corruption Indices Measure?” Unpublished paper, Harvard University and University of Houston, August 13, 2009.

Fisman, Raymond, and Roberta Gatti. 2002. “Decentralization and Corruption: Evidence Across Countries,” *Journal of Public Economics* Vol. 83, Nº 3: 325-45.

Galtung, Fredrik. 2006. “Measuring the Immeasurable: Boundaries and Functions of (Macro) Corruption Indices,” pp. 101-30, in Charles Sampford, Arthur Shacklock, Carmel Connors and Fredrik Galtung (eds.), *Measuring Corruption* (Aldershot, England; Burlington, VT: Ashgate).

Glaeser, Edward, Rafael LaPorta, Florencio Lopez-de-Silanes and Andrei Shleifer. 2004. “Do Institutions Cause Growth?” *Journal of Economic Growth* Vol. 9, Nº 3: 271-303.

Golden, Miriam A., and Lucio Picci. 2005. “Proposal for a New Measure of Corruption. Illustrated with Italian Data,” *Economics and Politics* Vol. 17, Nº 1: 37-75.

Hawken, Angela and Jonathan Kulick. 2006. “Public-employee Perceptions of Corruption in Post-transition Georgia,” Middle East and Central Asia Politics, Economics, and Society Conference, Salt Lake City, September 7-9, 2006.

Hawken, Angela and Gerardo L. Munck. 2008. “A Corruption Monitoring System for Afghanistan,” Paper prepared for the UNDP Accountability and Transparency (ACT) project, Kabul, Afghanistan, July 2008.

-------------- . 2009. “Measuring Corruption. A Critical Assessment and a Proposal,” pp. 71-106, in Anuradha K. Rajivan and Ramesh Gampat (eds.), *Perspectives on Corruption and Human Development* Vol. 1 (New Delhi: Macmillan India for UNDP).

Heidenheimer, Arnold J., and Michael Johnston (eds). 2002. *Political Corruption: Concepts and Contexts* 3rd. ed. (New Brunswick, NJ: Transaction Press).

Hsieh, Chang-Tai, and Enrico Moretti. 2006. “Did Iraq Cheat the United Nations? Underpricing, Bribes, and the Oil for Food Program,” *Quarterly Journal of Economics* Vol. 121, Nº 4: 1211-48.

Iqbal, Kazi, and Anwar Shah. 2008. “How Do Worldwide Governance Indicators Measure up?” Unpublished paper, World Bank, April 5, 2008. Available at: http://siteresources.worldbank.org/PSGLP/Resources/Howdoworldwidegovernanceindicatorsmeasureup.pdf

Kaufmann, Daniel, and Aart Kraay. 2008. "Governance Indicators: Where Are We, Where Should We Be Going?" *The World Bank Research Observer* Vol. 23, Nº 1: 1-30.

Kaufmann, Daniel, Aart Kraay and Massimo Mastruzzi. 1999. "Aggregating Governance Indicators," World Bank *Policy Research Working Paper* Nº 2195 (Washington, DC: The World Bank).

-------------- . 2004. "Governance Matters III: Governance Indicators for 1996, 1998, 2000, and 2002," *World Bank Economic Review* Vol. 18, Nº 2: 253-87.

-------------- . 2005. "Governance Matters IV: Governance Indicators for 1996-2004," World Bank *Policy Research Working Paper* Nº 3630 (Washington, DC: The World Bank).

-------------- . 2006. "Governance Matters V: Aggregate and Individual Governance Indicators for 1996–2005," World Bank *Policy Research Working Paper* Nº 4012 (Washington, DC: The World Bank).

-------------- . 2007a. "Growth and Governance: A Reply," *Journal of Politics* Vol. 69, Nº 2: 555–62.

-------------- . 2007b. "Growth and Governance: A Rejoinder," *Journal of Politics* Vol. 69, Nº 2: 570–72.

-------------- . 2007c. "The Worldwide Governance Indicators Project: Answering the Critics," World Bank *Policy Research Working Paper* Nº 4149 (Washington, DC: The World Bank).

-------------- . 2007d. "Governance Matters VI: Governance Indicators for 1996-2006," World Bank *Policy Research Working Paper* Nº 4280 (Washington, DC: The World Bank).

-------------- . 2008. "Governance Matters VII: Aggregate and Individual Governance Indicators, 1996-2007," World Bank *Policy Research Working Paper* Nº 4654 (Washington, DC: The World Bank).

Kaufmann, Daniel, Aart Kraay and Pablo Zoido-Lobatón. 2002. "Governance Matters II: Updated Indicators for 2000/01," World Bank *Policy Research Working Paper* Nº 2772 (Washington, DC: The World Bank).

Knack, Stephen. 2007. "Measuring Corruption A Critique of Indicators in Eastern Europe and Central Asia," *Journal of Public Policy* Vol. 27, Nº 3: 255-91.

Kurtz, Marcus J., and Andrew Schrank. 2007a. "Growth and Governance: Models, Measures, and Mechanisms," *Journal of Politics* Vol. 69, Nº 2: 538–54.

-------------- . 2007b. "Growth and Governance: A Defense," *Journal of Politics* Vol. 69, Nº 2: 563–69.

Lall, Sanjaya. 2001. "Competitiveness Indices and Developing Countries: An Economic Evaluation of the Global Competitiveness Report," *World Development* Vol. 29, Nº 9: 1501-25

Lambsdorff, Johann Graf. 1999. "The Transparency International Corruption Perceptions Index 1999. Framework Document," Transparency International. Available at: http://www.transparency.org/policy_research/surveys_indices/cpi

-------------- . 2006a. "Consequences and Causes of Corruption: What do We Know from a Cross-Section of Countries?" pp. 3-51, in Susan Rose-Ackerman (ed.), *International Handbook on the Economics of Corruption* (Cheltenham UK and Northhampton MA.: Edward Elgar).

-------------- . 2006b. "Measuring Corruption: The Validity and Precision of Subjective Indicators (CPI)," pp. 81-99, in Charles Sampford, Arthur Shacklock, Carmel Connors and Fredrik Galtung (eds.), *Measuring Corruption* (Aldershot, England; Burlington, VT: Ashgate).

-------------- . 2006c. “The Methodology of the Corruption Perceptions Index 2006**,”** Transparency International and University of Passau. Available at: http://www.transparency.org/policy_research/surveys_indices/cpi

-------------- . 2007. *The Institutional Economics of Corruption and Reform: Theory, Evidence and Policy* (Cambridge: Cambridge University Press).

-------------- . 2008. “The Methodology of the Corruption Perceptions Index 2008**,”** Transparency International and University of Passau. Available at: http://www.transparency.org/policy_research/surveys_indices/cpi

Langbein, Laura, and Stephen Knack. 2008. “The Worldwide Governance Indicators and Tautology: Causally Related Separable Concepts, Indicators of a Common Cause, or Both?” World Bank *Policy Research Working Paper* Nº 4669 (Washington, DC: The World Bank).

Mehlkop, Guido, Peter Graeff and Robert Neumann. 2008. “What do we Measure if we Measure Corruption? A Multitrait Multimethod Analysis,” Paper presented at the annual meeting of the American Sociological Association, Boston, MA. Available at: http://www.allacademic.com/meta/p240391_index.html

Mocan, Naci. 2008. “What Determines Corruption? International Evidence from Micro Data,” *Economic Inquiry* Vol. 46, Nº 4: 493-510.

Munck, Gerardo L. 2009. *Measuring Democracy: A Bridge Between Scholarship and Politics* (Baltimore, Md.: The Johns Hopkins University Press).

Olken, Benjamin A. 2007. “Monitoring Corruption: Evidence from a Field Experiment in Indonesia,” *Journal of Political Economy* Vol. 15, Nº 2: 200-49.

-------------- . 2009. “Corruption Perceptions vs. Corruption Reality,” *Journal of Public Economics* Vol. 93, Nº 7-8: 950-64.

Razafindrakoto, Mireille, and François Roubaud. 2006. “Are International Databases on Corruption Reliable? A Comparison of Expert Opinion Surveys and Household Surveys in sub-Saharan Africa*,” Document de Travail* 2006-17 (Paris: DIAL).

Reinikka, Ritva, and Jakob Svensson. 2004 “Local Capture: Evidence from a Central Government Transfer Program in Uganda,” *Quarterly Journal of Economics* Vol. 119, Nº 2: 679-706.

Sampford, Charles, Arthur Shacklock, Carmel Connors and Fredrik Galtung (eds.). 2006. *Measuring Corruption* (Aldershot, England; Burlington, VT: Ashgate).

Thomas, Melissa A. 2010. “What do the Worldwide Governance Indicators Measure?” *European Journal of Development Research* Vol. 22, Nº 1: 31-54.

Transparency International. 2009. “Corruption Perception Index (CPI).” Available at: www.transparency.org/policy_research/surveys_indices/cpi

Transparency International and Johann Graf Lambsdorff. 1997. “Transparency International publishes 1997 Corruption Perception Index,” Berlin. Available at: http://www.transparency.org/policy_research/surveys_indices/cpi

Treisman, Daniel. 2000. “The Causes of Corruption: A Cross-national Study,” *Journal of Public Economics* Vol. 76, Nº 3: 399-457.

-------------- . 2007. “What Have We Learned About the Causes of Corruption from Ten Years of Cross-National Empirical Research?” *Annual Review of Political Science* Vol. 10: 211-44.

Van de Walle, Steven. 2006. “The State of the World’s Bureaucracies,” *Journal of Comparative Policy Analysis* Vol. 8, Nº 4: 439-450

World Bank. 2005. "Country Policy and Institutional Assessments. 2005 Assessment Questionnaire." Available at: siteresources.worldbank.org/IDA/Resources/CPIA2005Questionnaire.pdf

-------------- . 2006. "Table 1: 2005 IDA Resource Allocation Index (IRAI)." Available at: http://web.worldbank.org/WBSITE/EXTERNAL/EXTABOUTUS/IDA/0,,contentMDK:21359477~menuPK:2626968~pagePK:51236175~piPK:437394~theSitePK:73154,00.html

-------------- . 2007a. "WGI: Worldwide Governance Indicators Country Snapshot. Control of Corruption." Available at: http://info.worldbank.org/governance/kkz2005/tables.asp

-------------- . 2007b. "WGI: Worldwide Governance Indicators Country Snapshot Comparison for Control of Corruption across a Number of Countries." Available at: http://info.worldbank.org/governance/kkz2005/mc_region.asp

-------------- . 2009a. "Governance Matters 2008. Worldwide Governance Indicators, 1996-2007," Available at: http://info.worldbank.org/governance/wgi/sc_country.asp

-------------- . 2009b. "Governance Matters 2008. WGI Weights," Available at: http://info.worldbank.org/governance/wgi/sc_country.asp