

What Do the Worldwide Governance Indicators Measure?

M.A. Thomas*

Associate Professor of International Development
The Paul H. Nitze School of Advanced International Studies
The Johns Hopkins University
ma.thomas@jhu.edu
Draft of October, 2006

Abstract

Since the 1990s, researchers and donors have focused more closely on the importance of “good governance” as a determinant of development and as a development objective in itself. The emphasis on good governance has created a demand for indicators to measure the quality of governance for both research and aid targeting. The Worldwide Governance Indicators, which rank countries on six aspects of “good governance” are first generation aggregate indicators that are relied upon by both researchers and policymakers. This paper considers the “construct validity” of these indicators -- whether the indicators measure what they purport to measure. The paper concludes that while there are reasonable questions about the construct validity of the indicators, no evidence of construct validity has been offered. Moreover, given the methodology employed, evidence of construct validity is not likely to be forthcoming. To date, the indicators stand as an elaborate untested hypothesis about the nature of governance. Reliance upon them for any purpose is premature.

Since the 1990s, development researchers and practitioners have focused on “good governance” as both a means of achieving development and a development objective in itself. The World Bank has defined “good governance” as “epitomized by predictable, open and enlightened policy making; a bureaucracy imbued with a professional ethos; an executive arm of government accountable for its actions; and a strong civil society participating in public affairs; and all behaving under the rule of law.” (World Bank, 1994)

In response to the growing demand for measures of the quality of governance, a number of aggregate governance indicators have been produced, such as the World Bank’s Worldwide Governance Indicators (“WGI”).¹ The WGI rank countries with respect to six aspects of good governance: Voice and Accountability, Political Stability and Violence, Government Effectiveness, Rule of Law, and Control of Corruption. These

* © 2006 Melissa Thomas. The author would like to thank Ray Koopman of Simon Fraser University, Francis Fukuyama of The Johns Hopkins University and David Armor of George Mason University for their very helpful comments and insights, and her research assistant, Scott Merrill, for his able assistance.

¹ The indicators have had various names over the years, but this paper uses the name used on the World Bank’s web site and most recently by the authors.

indicators have been used by researchers as explanatory variables and by United States policymakers to allocate enormous aid packages.

The authors of the indicators have tried to draw attention to the large standard errors of the estimates – caveats that have been largely ignored. But researchers and policymakers should not only be concerned with large standard errors. Before they rely on these indicators, they should ask much more fundamental questions. What do these indicators measure, if anything, and how would we know? Section I describes the WGI and their use. Section II discusses the process of building indicators of abstract concepts and the presentation of evidence of validity. Section III describes the methodology used to estimate the WGI. Section IV analyzes the methodology in light of the need for evidence of construct validity. Section V concludes that the indicators stand as an elaborate and unsupported hypothesis. Accordingly, reliance on the indicators is premature.

I. The Worldwide Governance Indicators

Since the 1990s, the World Bank as well as researchers and other donors have focused more closely on the role of good governance in development. The economic historian Douglass North, among others, drew attention to the institutional prerequisites for economic growth. (North et al., 1996) An influential report by World Bank researchers claimed that foreign aid could only be used effectively in well-governed countries, although their results are now disputed. (Burnside et al., 2004; Easterly et al., 2004; World Bank, 1998) Former President of the World Bank, James Wolfensohn brought the issue of corruption into the mainstream of the Bank’s dialogue with its borrowers. (Wolfensohn, 1999) Current President Paul Wolfowitz has committed to deepening the Bank’s work on governance and corruption. (World Bank, 2006a)

In order to test claims about the importance of good governance, or to implement policies that aim either to strengthen governance or target aid to well-governed countries, measurements of the quality of governance are needed. While some data are available, they are problematic because they are not always good quality and they are not comparable. Data come from expert assessments, polls of experts, and surveys of government officials, businesses and households. A few sources aim at global coverage, but the coverage of most sources is much more limited. The surveys and polls from various sources do not share a common methodology, definition of terms, set of questions or measurement scale of responses.

World Bank researchers attempted to address these problems by developing aggregate governance indicators that draw from many available sources. The 1999 papers “Aggregating Governance” and “Governance Matters” introduced the Worldwide Governance Indicators. (Kaufmann et al., 1999a; Kaufmann et al., 1999b) The indicators are defined to correspond to what the authors consider to be “fundamental governance concepts.” (Kaufmann et al., 1999b) Most recently, the indicators were defined as:

1. *Voice and Accountability* (VA), the extent to which a country's citizens are able to participate in selecting their government, as well as freedom of expression, freedom of association, and free media.
2. *Political stability and absence of violence* (PV), perceptions of the likelihood that the government will be destabilized or overthrown by unconstitutional or violent means, including political violence or terrorism.
3. *Government effectiveness* (GE), the quality of public services, the quality of the civil service and the degree of its independence from political pressures, the quality of policy formulation and implementation, and the credibility of the government's commitment to such policies.
4. *Regulatory quality* (RQ), the ability of the government to formulate and implement sound policies and regulations that permit and promote private sector development.
5. *Rule of law* (RL), the extent to which agents have confidence in and abide by the rules of society, and in particular the quality of contract enforcement, the police, and the courts, as well as the likelihood of crime and violence.
6. *Control of corruption* (CC), the extent to which public power is exercised for private gain, including both petty and grand forms of corruption, as well as "capture" of the state by elites and private interests.

(Kaufmann et al., 2006a) The indicators are not absolute measures of governance, but are measures of a country's relative rank with respect to that indicator. The indicators have been produced for 1996, 1998, 2000, 2002, 2004 and 2005 with global coverage, albeit with some missing values. Six papers have been written presenting the data sets. Most are published as World Bank Working Papers, but one has been published in a peer-reviewed journal. (See Kaufmann et al., 2004, 2005, 2006a; Kaufmann et al., 1999a, 2002; Kaufmann et al., 1999b)

The WGI cover 213 countries and territories and are based on several hundred variables produced by 25 different sources, including commercial data providers. (Kaufmann et al., 2006a) While the authors have drawn explicit attention to the large standard errors associated with the governance estimates, they argue that the methodology employed for developing the indicators has two important strengths. First, the aggregation methodology makes the WGI more informative than any individual data source. Second, it allows calculation of the margins of error of the estimated indicators. (Kaufmann et al., 2006a)

The global coverage of the indicators and the claim that they are the most precise measure of governance make the indicators attractive to researchers and policymakers alike. In "Governance Matters," Kaufmann et al. use the indicators to find "a strong

positive causal relationship from improved governance to better development outcomes.” (Kaufmann et al., 1999b) A number of other studies have also used the indicators as explanatory variables. (See, e.g., Andres, 2006; Apodaca, 2004; Clarke et al., 2006; Das et al., 2006; Hart et al., 2005; Jung, 2006; Liu et al., 2006; Llamazares, 2005; Neumayer, 2002)

Policymakers are also employing the indicators. The Millennium Challenge Account (“MCA”) was launched by the Bush administration to target large grants of foreign aid to countries that are well-governed compared to their income peers. To make the determination of eligibility, the MCA relies on sixteen third-party indicators of the quality of governance divided into three categories: Ruling Justly (six indicators), Encouraging Economic Freedom (six indicators) and Investing in People (four indicators). To qualify for funding, a country “must perform above the median in relation to its peers on at least half of the indicators in each of the three policy categories and above the median on the corruption indicator.” (Millennium Challenge Account, 2006) Of the sixteen indicators, five are WGI. WGI make up four of the six indicators in the policy category, “Ruling Justly.” Finally, the Governance Research Indicator “Controlling Corruption” is the corruption indicator used to disqualify any country that falls below the median. The indicator methodology is the “predominant method” for determining eligibility although discretion also plays a role. (Millennium Challenge Account, 2006)

In addition to the U.S. government, according to the World Bank, “[o]ther donor governments, such as the Netherlands, also rely on the Worldwide Governance Indicators to monitor the quality of governance in aid recipient countries. Risk rating agencies as well as [non-governmental organizations] also use them.” (World Bank, 2006b)

The authors have taken care to point out the large margins of error of their estimates. Kaufmann et al. report that of the 70 countries identified as potential MCA beneficiaries for the 2005 fiscal year, about 40 countries could be placed above or below the median with 90% confidence; the remaining 30 fell into a “zone of uncertainty.” Relaxing the confidence level to 75% still left twenty countries in this zone. (Kaufmann et al., 2005)

The increasing reliance on the WGI has prompted other scholars to take a closer look at the construction of the indicators, the claims made by the authors regarding the indicators, and the ways in which third parties make use of them. (See Arndt et al., 2006; Knack, 2006) In addressing these critiques, Kaufmann, Kraay and Mastruzzi group them into several categories: concerns about the comparability of the indicators across countries and across time; concerns about bias in expert polls or in particular sources; and concerns about the independence of the different data sources and the consequences for the aggregate indicators. The authors have addressed these issues at length. (Kaufmann et al., 2006b)

While issues of standard error and comparability are important ones, there are some more fundamental questions that both researchers and policymakers should ask before they rely on the WGI or any other measure of governance. What do these indicators measure?

What does it mean to measure an abstract concept like “government effectiveness” or “rule of law”? Do these indicators measure anything at all? And how would we know?

II. Construct Validity

Social scientists have been developing measures of abstract concepts ever since psychologists attempted to define and measure intelligence in the early twentieth century. (Bartholomew, 1995; Williams et al., 2003) Quantitative measures of abstract concepts have been used in psychometrics, to measure psychological attributes and aptitudes (see, e.g., Patterson, 1990); in education, to measure educational achievement and ability (see, e.g., Forsythe et al., 1986); in public health, to measure variables such as cognitive health (see, e.g., Wallace et al., 1995); in public administration, to measure public service motivation (see, e.g., Perry, 1996); in economics, to measure contingent valuation and averting costs (see, e.g., Laughland et al., 1996); and in political science to measure political perceptions and attitudes, as well as political variables such as democracy (see, e.g., Elkins, 2000; Faber, 1987).

Building scales to measure inherently unobservable general concepts, or “constructs,” involves special considerations. First, it requires a mapping between a theory about the construct and a specific definition of the construct that is a description of the thing to be measured. Second, it requires a mapping between the description and a specific operationalization of that idea, a model based on observable variables that is used to derive a measure of the construct. Figure 1 illustrates these relationships, while making the point that there may be any number of equivalent operationalizations of the same construct. Lastly, it requires predictions about how the construct (and therefore the proposed measurement of the construct) relates to other observables. These predictions both provide a means to check the correctness of the choices made in operationalization and an explanation of why we would care about the construct at all.

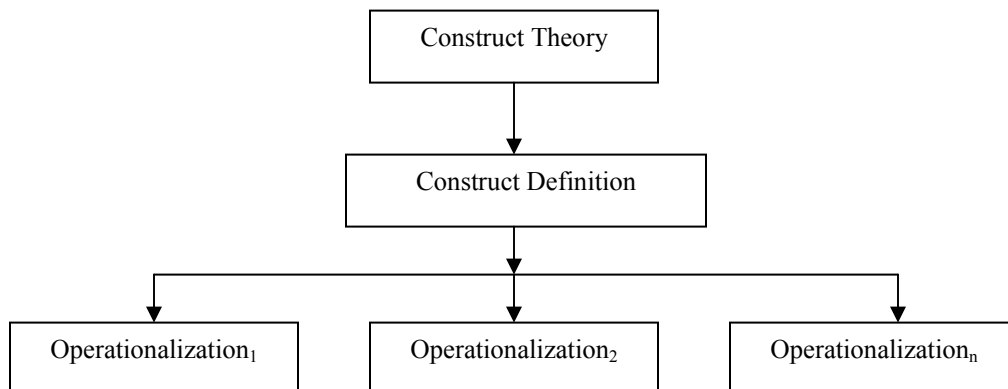


Figure 1. Measuring abstract concepts

To be scientifically admissible, an indicator that results from a specific operationalization must be shown to be “valid,” which means that it measures what it purports to measure. (Carmines et al., 1979) As Adcock and Collier caution,

it is essential to view the interpretations of scores in relation to systematized concepts as falsifiable claims. . . . Scholars should treat these claims just as they would any causal hypothesis, that is, as tentative statements that require supporting evidence. Validity assessment is the search for this evidence.

(Adcock et al., 2001).

Errors in the mapping process can lead to invalid measures. The literature on construct validity has identified a number of different types of validity that correspond with specific errors:

- “Content validity” or “face validity” is concerned with whether the proposed operationalization captures the entire domain of a construct and includes nothing extraneous. Carmines and Zeller offer the example of a test that purports to measure the capacity to perform arithmetic operations, but contains only addition problems, and no subtraction, multiplication or division problems. (Carmines et al., 1979)
- “Convergent validity” and “discriminant validity” are concerned, respectively, with the extent to which measurements that should in theory be correlated, are correlated, and the extent to which measurements that should in theory not be correlated are sufficiently uncorrelated.
- “Criterion validity” is concerned with whether a measure that is supposed to be a proxy measure of an observable phenomenon is correlated with direct measures of that observable.

(See Adcock et al., 2001; Campbell et al., 1959; Carmines et al., 1979)

Convergent and discriminant validity are of particular importance in the development of measures of theoretical constructs. Cambell and Fiske coined the terms arguing that

[A]ny conceptual formulation of trait will usually include implicitly the proposition that this trait is a response tendency which can be observed under more than one experimental condition and that this trait can be meaningfully differentiated from other traits. The testing of these two propositions must be prior to the testing of other propositions to prevent the acceptance of erroneous conclusion.

(Campbell et al., 1959) The measurement of a construct has convergent validity if it is correlated with measurements of the same trait by different methods, which provides

evidence of the existence of a common trait. The measurement of a construct has discriminant validity if it is not correlated with other measurements from which it is expected to differ, which provides evidence that it is a distinct trait.

Unlike measures that are simple counts of events or objects, a measure of a theoretical construct is necessarily embedded in theory, which is what imbues the measure with meaning. First, the researcher draws on theory to derive a definition of the construct and to make clear how the new construct is distinct from existing constructs. (Campbell et al., 1959) This definition is not a one-line description but a “fleshed-out account” of the concept that is within the matrix of meanings usually assigned to the construct by scholars. (Adcock et al., 2001) Second, the researcher draws on theory to describe how the construct, assuming that it is correctly measured, should relate to other observables. “For hypothetical constructs, there is no good way to determine whether a measure reflects the construct validly, except to examine whether scores on the measure conform to a theory, of which the target construct is a part.” (Smith, 2005) (See also Carmines et al., 1979; Cronbach et al., 1955) It is this network of predicted relationships with observables, named the “nomological network” by Cronbach and Meehl, that anchors the construct in the world of science. (Cronbach et al., 1955) An atheoretical measure of a theoretical construct is a contradiction in terms.

The proposed construct, its measure, and the theory from which they spring, gain scientific acceptance through a public iterative process of evaluation and theory revision. Accordingly, “each component of a research program, or each component of theory derivation, hypothesis formation, and empirical test, must be open to criticism.” (Smith, 2005)

A number of techniques and tools have been brought to bear on the process of testing and providing evidence of construct validity. Smith provides a recent survey of quantitative approaches. (Smith, 2005) To check for convergent and discriminate validity, Campbell and Fiske proposed an examination of the correlation matrix showing correlations among measures of more than one trait, measured by independent methods. (Campbell et al., 1959) Using this Multitrait-Multimethod Matrix (“MTMM”), researchers check that measures of the same trait are more highly correlated with each other than with measures of different traits. Other researchers have used structural equation modeling to isolate different sources of variance. (See, e.g., Bollen, 1989, 1993; Eid et al., 2003; Hammond et al., 1986) Others have proposed indices that measure the goodness-of-fit of the proposed operational model. (Westen et al., 2003)

The first question that should occupy potential users of any governance indicator is not the size of the margins of error, but whether the indicators are valid measurements of what they purport to measure.² A purported measure of a theoretical construct must be

² A second necessary property of a measure of a construct is reliability. Validity and reliability are usually discussed together. A measure is “reliable” “to the extent to which an experiment, test, or any measuring procedure yields the same results on repeated trials.”

Carmines, E. G. and R. A. Zeller (1979) *Reliability and Validity Assessment*. Beverly Hills: Sage Publications. The WGI will be reliable to the extent that the underlying individual indicators are reliable.

shown to be valid, both in its conceptualization and its operationalization, by exploring predicted relationships with other observable variables.

III. Method of the Construction of the WGI

The estimates of governance produced by the WGI Project are the result of up to three levels of aggregation of underlying survey responses or expert assessments. The methodology is described most fully in (Kaufmann et al., 1999a) and nicely summarized in (Arndt et al., 2006). Figure 2 illustrates the model underlying the WGI. Kaufmann, Kraay and Mastruzzi first identify interesting variables produced by third party sources. Much of this data is proprietary and at least one source is confidential.³ While some of the indicators are responses to survey questions or to questions asked of experts, some of these indicators are themselves aggregates produced by third parties and the methodology for their construction – even their definitions – are not publicly available.

The indicators are rescaled for comparability. The authors then sort the indicators⁴ into six clusters, each of which corresponds to what the authors believe is a “fundamental concept” of governance.⁵ Each indicator is assigned to only one cluster. “This classification of indicators into clusters corresponding to this definition of governance is not intended to be definitive. Rather, it simply reflects our views of what constitutes a consistent and useful organization of the data that is concordant with prevailing notions of governance.” (Kaufmann et al., 2005) In order to eliminate correlated errors among the indicators arising from source-specific perception errors, indicators from the same data source are averaged and this simple average (referred to in this paper as a “source variable”) is used in place of the indicators themselves.

Each source variable in a cluster is assumed to be a “noisy” observation of the corresponding unobservable governance construct. The problem is then one of signal extraction:

We use an extension of the standard unobserved components model which [sic] expresses the observed data in each cluster as a linear function of the

Accordingly, the reliability of the WGI is beyond the control of the authors and will not be discussed here except to note that the authors recognize the importance of measurement error and have paid considerable attention to minimizing the impact of such error.

³ Commercial data providers such as Business Environment Risk Intelligence sell their data. Until 2005, the World Bank’s Country Policy Institutional Assessments were confidential and not available even to well-financed researchers.

⁴ Kaufmann, Kraay and Mastruzzi use the word “indicator” sometimes to refer to the underlying individual indicators, sometimes to the average of these indicators by source, and sometimes to the Worldwide Governance Indicators. In this paper, the individual indicators will be referred to as “indicators,” the first aggregation of these indicators by source as “source variables,” and the final governance indicators as “governance indicators” or “estimates.”

⁵ The indicators, sources and assignment of indicators to clusters is set out in Appendix B of “Governance Matters V” and repeated for convenience in Appendix A of this paper.

Kaufmann, D., A. Kraay and M. Mastruzzi (2006a) 'Governance Matters V: Aggregate and Individual Governance Indicators for 1996-2005'. Washington, D.C.: World Bank.

unobserved common component of governance, plus a disturbance term capturing perception errors and/or sampling variation in each indicator. In particular, we assume that we can write the observed score of a country j on indicator k , $y(j,k)$, as a linear function of unobserved governance, $g(j)$, and a disturbance term, $\varepsilon(j,k)$, as follows:

$$y(j,k) = \alpha(k) + \beta(k) \cdot (g(j) + \varepsilon(j,k))$$

where $\alpha(k)$ and $\beta(k)$ are unknown parameters which [sic] map unobserved governance $g(j)$ into the observed data $y(j,k)$.

(Kaufmann et al., 2005) Accordingly, each source variable is assumed to be a function of one and only variable, namely the unobserved governance component for that cluster. The relationship between the source variable and the unobserved governance component is assumed to be linear.

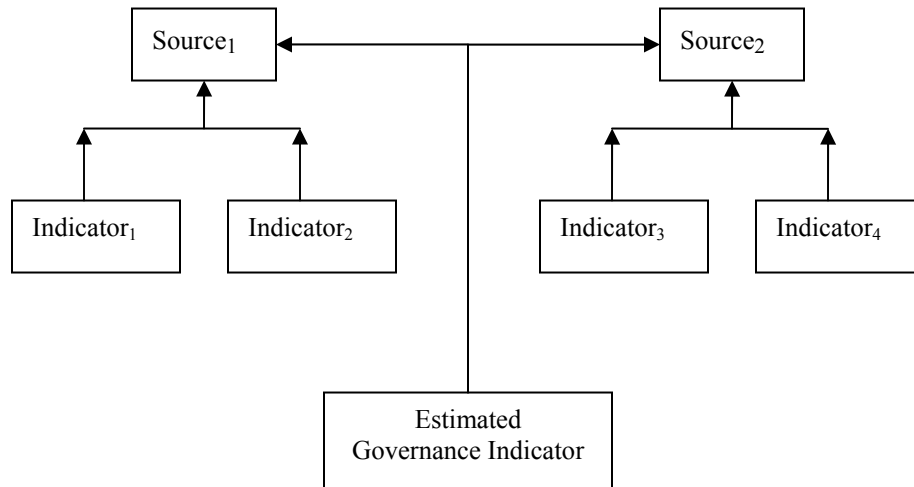


Figure 2. Model of governance underlying the WGI.

The model depends on a number of important assumptions regarding the error terms, which capture both measurement errors due to perception errors or sampling errors and errors due to imperfect relationships between the indicator and the governance concept. (Kaufmann et al., 2005) The error terms are assumed to have a zero mean and to be independent across source variables and countries. That is

$$E[\varepsilon(j,k)\varepsilon(j',k')] = 0 \text{ if } j \neq j' \text{ or } k \neq k'$$

(Kaufmann et al., 1999a) The variance of the error terms is assumed to differ across indicators but to be the same across countries. In other words,

$$E[\varepsilon(j,k)^2] = \sigma_\varepsilon^2(k)$$

(Kaufmann et al., 1999a) Governance and error terms are assumed to be jointly normally distributed. (Kaufmann et al., 1999a) Finally, the distribution of unobserved governance and the mean of world governance are assumed to be the same in every period; in other words, that world governance does not ever get better or worse over time.

To produce estimated governance indicators for a particular governance component for a country, the authors produce a weighted average of the governance estimates for the source variables in the cluster for that country. The weights are inversely proportional to the variance of the error term of the source. (Kaufmann et al., 2005) According to the authors, this method “rewards conformity”, in the sense that indicators that are highly correlated will have low estimated variances and hence will be perceived as more precise.” (Kaufmann et al., 1999a)

IV. Are the WGI Valid Measures of Governance?

The methodology as presented raises several concerns. The first is that the constructs themselves are poorly defined and may be meaningless. The second is the likelihood of model specification errors and the lack of evidence of model fit. The last is that no evidence for construct validity has been presented; indeed, given the methodological choices, it is doubtful that it could be.

1. Construct Definition: The Risk of Meaninglessness

The WGI purport to measure fundamental concepts of governance, but these concepts are neither well-defined nor rooted in theory. It is notable that none of the papers presenting the WGI contains any reference to the theoretical literature on governance. The terms “government effectiveness” and “regulatory quality” are not associated with established theoretical literatures, and no alternate theory is articulated. On the other hand, terms such as “voice,” “accountability,” and “rule of law” are concepts featured in existing theoretical literatures, but the definitions of the WGI neither refer to nor correspond with these literatures. For example, the indicator “Voice and Accountability” is defined as “the extent to which a country’s citizens are able to participate in selecting their government, as well as freedom of expression, freedom of association and free media.” (Kaufmann et al., 2005) In political science, the term “voice” derives from Hirschman’s classic work, “Exit, Voice and Loyalty.” Hirschman gives the following definition of “voice”:

any attempt at all to change, rather than escape from, an objectionable state of affairs, whether through individual or collective petition to the management directly in charge, through appeal to a higher authority with the intention of forcing a change in management, or through various types of actions and protests including those that are meant to mobilize public opinion.

(Hirschman, 1970) Accordingly, “voice” is not synonymous with accountability, freedom to select government, or other political freedoms. Nor are there well-known or well-documented relationships between them.

The definition of the term “rule of law” is the subject of wide debate in a post-natural law world. According to Fallon, modern definitions tend to refer to five constituent elements. These are (1) “the capacity of legal rules, standards or principles to guide people in the conduct of their affairs”; (2) efficacy; (3) stability; (4) “the supremacy of legal authority” for both citizens and government actors; and (5) the availability of impartial institutions of enforcement. (Fallon, 1997) The definition of the Rule of Law governance indicator - - “the extent to which agents have confidence in and abide by the rules of society, and in particular the quality of contract enforcement, the police and the courts, as well as the likelihood of crime and violence” -- captures some elements of this definition, includes elements not traditionally incorporated in the concept, and omits others. As such, this construct does not have “content validity” – it does not map to a definition of the Rule of Law.

That the names of the governance indicators are divorced from the literatures dealing with constructs of the same names does not mean that the governance indicators are wrong or useless. They may simply be new constructs that are misnamed. To introduce the new construct, the authors would have to explain the nature and importance of the new construct, the distinctions between the new construct and existing constructs, and the predicted relationship of the new construct to other variables of interest before attempting an operationalization. (It might also help to find a less misleading name.)

An examination of the underlying indicators suggests that the WGI construct definitions are merely summary descriptions of the indicators in the cluster. The authors explain the clustering as an expression of personal ideas of governance, but these ideas are not articulated. (Kaufmann et al., 2005) If this is the case, then the underlying construct has not been defined at all and there is no way to check the validity of these constructs or to explain why we would care about them. No meaningful predictions can be made regarding the relationship between undefined constructs and observable variables.

2. Model Specification: The Risk of Error

Once the construct is well defined, the researcher must specify an operationalization of the construct in terms of observables. An econometric model is specified by choosing a dependent variable, independent variables, a functional form that relates the two, and a

type of stochastic error term. Studemund's introductory economics text includes this caution: "Of all the kinds of mistakes made in applied regression analysis, specification error is usually the most disastrous to the validity of the estimated equation." (Studemund, 1997) Accordingly, it is critical that researchers explain and justify the assumptions that underlie their models, which are presumed to be theoretically motivated. This section discusses two types of possible specification errors: errors in clustering, and omitted variables.

a. Clustering

The process of clustering indicators involves two important considerations. The first is the determination of the number of clusters that best represents the underlying data. The authors assume that the underlying data is best represented by six unobserved governance constructs. Why six? Why not nine, or two? Conflating indicators determined by different unobserved governance components would make the estimates meaningless. Separating indicators determined by the same component into multiple clusters unnecessarily would result in unnecessarily large margins of error.

How can one tell if the data reflect one or more unobserved components of governance? There are no accepted tests for establishing the unidimensionality of data. (See, e.g., Dunn et al., 1973; Hattie, 1984; See, e.g., Nandakuma et al., 1993; Segars, 1997) Researchers typically use factor analysis to try to break a body of data into distinct components. The process is one of trying for the best fit; however there is no suggestion that the development of the WGI involved an exploration of the dimensionality of the underlying data; instead it was assumed.

Table 1 shows the pairwise correlation among the governance indicators for the data for the year 2005. In some cases, the correlations are so high that they raise the question of whether the indicators are measuring distinct constructs – or, put another way, whether they have discriminant validity. The Control of Corruption and the Rule of Law indicators, the Control of Corruption and the Government Effectiveness indicators, and the Rule of Law and Government Effectiveness indicators are correlated at 0.95, while Government Effectiveness and Regulatory Quality indicators are correlated at 0.96.

	Voice and Accountability	Political Stability and Violence	Government Effectiveness	Regulatory Quality	Rule of Law	Control of Corruption
Voice and Accountability	1.00					
Political Stability and Violence	0.70	1.00				
Government Effectiveness	0.79	0.75	1.00			
Regulatory Quality	0.80	0.72	0.96	1.00		
Rule of Law	0.78	0.81	0.95	0.91	1.00	
Control of Corruption	0.74	0.75	0.95	0.90	0.95	1.00

Table 1. Correlation matrix for the 2005 WGI (Kaufmann et al., 2006a)

Assuming that researchers have correctly identified the number of underlying governance components in a data set, the next problem is to ensure that each indicator is assigned to the proper cluster. Assigning indicators into the wrong cluster could obscure existing relationships or create spurious ones. Because the construct definitions lack specificity, the rationale for the assignment of indicators to clusters is unclear. Moreover, the names, definitions and clustering are inconsistent. If “Voice and Accountability” measures “the extent to which a country’s citizens are able to participate in selecting their government, as well as freedom of expression, freedom of association and free media,” what is the hypothesized relevance of the indicator “When deciding upon policies and contracts, Government officials favor well-connected firms”? Would this variable belong in the “Control of Corruption” cluster? What is the hypothesized relevance of the indicator “Institutional Stability” and what does it mean? Would this variable be better put in the “Government Effectiveness” cluster?

b. Assumptions of Independence

The WGI methodology combines indicators from the same source into a source variable; then the source variable is regressed on the unobserved governance component. The model assumes that errors are uncorrelated.

From inception, the authors have been concerned about reducing the possibility of correlated errors. This is because correlated error terms would increase the margins of error of the governance estimates. Moreover, because highly correlated sources are more heavily weighted, correlated errors could also result in misranked countries. (Kaufmann et al., 1999a) The aggregation of all indicators from the same source into a single source variable was intended to reduce the possibility of source-specific correlated errors. The most recent paper on the indicators addresses in detail a criticism that in particular, expert polls are likely to be correlated. The authors explore the impact of such correlated errors

on the standard errors of the indicators, and conclude that, even if such polls are correlated, the impact would be to raise the standard errors by about 10 percent. (See Kaufmann et al., 2006a, 2006b)

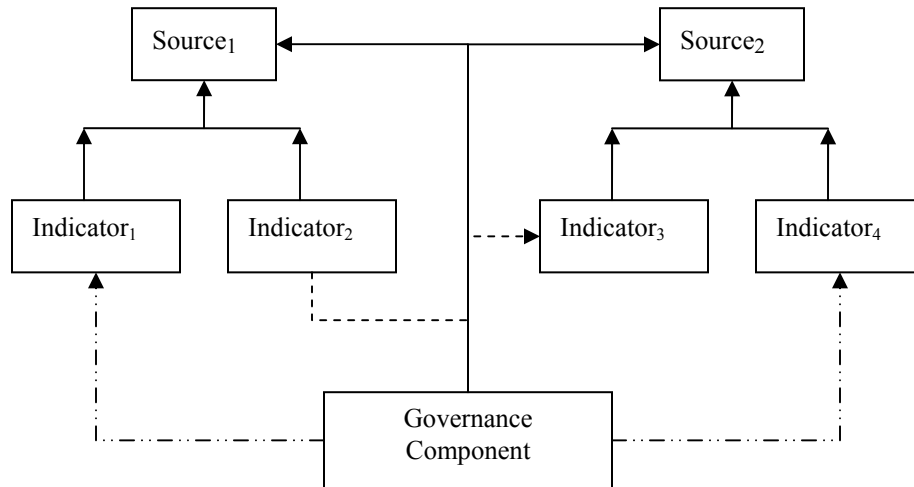


Figure 3. The dashed arrows illustrate possible functional relationships that would result in correlated error terms between the sources variables.

If there were functional relationships among indicators from different sources, or simultaneity issues between indicators and unobserved governance, source variables would also have correlated error terms.⁶ (See Figure 3). But specification errors of this type could not only increase the standard errors of the WGI, or result in misranked countries, but would also cause omitted variable bias, leading to inconsistent parameter estimates.⁷ (White, 1982) Errors could cause the estimates to shift as the underlying data sample changes, making year-to-year changes in country rankings meaningless.

An examination of the indicators in each cluster (see Appendix A) shows a number of theoretically plausible functional relationships among indicators from different sources. In the “Voice and Accountability” cluster, for example, indicators are included for “Freedom of the press” (Freedom House) and “Freedom of association” (Economist Intelligence Unit) on the one hand, and “Democratic accountability” (Political Risk

⁶ If an omitted variable affects an indicator, then its effect is captured in the error term. If the same omitted variable affects more than one indicator, then a rise in the variable will cause the error terms of the indicators to rise as well.

⁷ A consistent estimator is one that has a greater probability of converging on the population parameter the greater the sample size.

Wooldridge, J. M. (2003) *Introductory Economics: A Modern Approach*. N.p.: South-Western. An inconsistent estimator does not; increasing the sample size is no guarantee of improved accuracy.

Index) on the other. Is it likely that there is no functional relationship between the former and the latter except a shared common governance element? In the “Government Effectiveness” cluster, indicators are included for “Public spending composition” (Global Competitiveness Survey) on the one hand, and “Management of external debt” (CPIA) and “Government handling of education” (Afrobarometer Survey) on the other. Is it plausible that the management of public debt and the government’s handling of education are unrelated to the composition of public spending except through a shared common governance element? Indeed, some indicators have been included in clusters not because they are measures of the construct, but because of presumed functional relationships with other indicators in the cluster. For example, the definition of the “Extremism” indicator in the Political Stability and Violence cluster states that it is important *because* of its “close association with violence,” while indicators of levels of violence are also included. Any indicator included because it is hypothesized to cause or result from an indicator from another source presents the problem of functional relationships leading to correlated errors among the source variables as well as the problem of omitted variable bias.

In addition, there might be functional relationships among governance components. If any governance indicator is a function of another governance indicator, the indicators in one cluster would have functional relationships with indicators in another cluster. The indicator “Control of Corruption” measures “the extent to which public power is exercised for private gain, including both petty and grand forms of corruption, as well as ‘capture’ of the state by elites and private interests.” But how could a government with poor control of corruption score high on “Government Effectiveness,” defined to measure “the quality of public services, the quality of the civil service and the degree of its independence from political pressures, the quality of policy formulation and implementation, and the credibility of the government’s commitment to such policies”? Similarly, corrupt acts are for the most part criminal; accordingly, wouldn’t there be a relationship between corruption levels and the rule of law?

Yet another problem is presented where a governance construct is composed of constituent elements, as is the case in the theory regarding the rule of law. A measure of a constituent element would be an independent variable; the unobservable rule of law would be the dependent variable. The model of the WGI would then misspecify the direction of causality with respect to those elements.

The WGI stands as an elaborate untested hypothesis about governance involving hundreds of variables. While researchers are usually urged to be parsimonious in the development of models and theories, the WGI rises in complexity with the addition of each new governance variable, which is assumed to be independent of previously included variables from other sources. The Voice and Accountability governance indicator, for example, is derived from 35 underlying indicators from sixteen sources. As more variables are included, the complexity of the model increases along with the risk of a specification error. Some of the assumptions appear to be improbable on their face. What evidence supports the WGI hypothesis?

3. Lack of Construct Validity: Unsubstantiated Claims

No evidence of construct validity or of goodness of fit of the WGI model has been offered. Indeed, evidence of construct validity is likely to be very difficult to produce, for several reasons. The first is that any showing that a hypothetical construct has been meaningfully operationalized is an exercise rooted in theory about the construct. Most critically, researchers need to be able to make predictions derived from the theory and test their operationalization against those predictions. But while there is a body of theory about governance, and about constructs such as “voice,” “accountability” and the “rule of law,” the WGI Project makes no reference to this or any other theory. No predictions have been made for the relationship of these constructs to observables, and therefore no evidence of convergent or discriminate validity can be advanced.

A second reason why it may be difficult to produce evidence of construct validity is the decision to encompass as much available data as possible in the estimation. This approach leaves few independent sources that could be used either for comparison or model testing.

A final reason why it may be difficult to produce evidence of construct validity is because the raw data underlying the governance indicators is not publicly available. Moreover, this means that it is impossible for other researchers to critique, improve or build on the indicators in the iterative process of theory justification and refinement. Even the definitions and the method of collection of the underlying indicators are not in the public domain. Although in September 2006 the World Bank announced the publication of “the individual measures of governance from 28 of the 31 data sources” what was released were the aggregate source variables rather than the underlying indicators.

V. Conclusion

Since the 1950s, public evidence of construct validity has been required for proposed measurements of abstract constructs. Absent the rigorous examination of the inferences involved in creating a measurement, authors such as Underwood and Campbell and Fiske recognized “the danger . . . that the investigator will fall into the trap of thinking that because he went from an artistic or literary conception . . . to the construction of items for a scale to measure it, he has validated his artistic conception.” (Campbell et al., 1959)

Developing a meaningful measurement of a construct is an iterative process that involves a theoretical specification of the construct and its relationship to observable variables, model testing as against predictions, and refinement. The process of accumulating evidence to support a hypothetical measure is one that involves the scientific community, and, as such, an investigator must provide evidence of construct validity and make available the raw data, models and results that would allow the community to make independent judgments. In the words of Cronbach and Meehl, “Defending a claim of construct validity is a major task, not to be satisfied by a discourse without data. . . . A

claim is unsubstantiated unless the evidence for the claim is public, so that other scientists may review the evidence, criticize the conclusions, and offer alternative interpretations.” (Cronbach et al., 1955)

The WGI represent a complex atheoretical and as yet poorly articulated hypothesis for which no evidence has been advanced. Moreover, the data upon which they are based are not available to the research community to allow evaluation, critique or refinement. Despite the evident demand for global measures of the quality of governance, the work is still in its infancy and usage is premature. For researchers, results obtained using the indicators are uninterpretable and as such should not survive peer review. For policymakers, reliance on the indicators is arbitrary.

This paper has focused on the WGI because of rising third party use and in particular, the reliance on the WGI for important policy decisions. However, the concerns raised here about the WGI apply equally to other current aggregate governance indicators. Of this group, the WGI are notable because of the authors’ efforts to document and make public the methodology for producing the indicators, to employ a statistically sound approach, to seek precision, and to both calculate and emphasize the importance of margins of error.

Unfortunately, meaningful measures of governance require as a prerequisite specific definitions of governance that draw from available theory. Developing such definitions is technically challenging, but it is made even more difficult because in the foreign policy arena the definition of good governance is highly politicized. A liberal democracy and an authoritarian dictatorship can both agree on the importance of the rule of law, provided that the former means by the term “a state constrained by rules” and the latter means “citizen obedience to government edicts.” The consensus on the importance of good governance may not survive if it requires agreement on the details, and measures derived from specific and idiosyncratic definitions may not find wide audiences.

Appendix: Components of Aggregate Governance Indicators 2005

These tables are reproduced with minor revisions of spelling, grammar and capitalization from (Kaufmann et al., 2006a).

Table A. Indicators for Voice and Accountability

Voice and Accountability	
Representative Sources	
Source	Concept Measured
EIU	Orderly transfers Vested interests Accountability of public officials Human rights Freedom of association
FRH	Civil liberties: Freedom of speech, of assembly and demonstration, of religion, equal opportunity, of excessive governmental intervention Political rights: free and fair elections, representative legislative, free vote, political parties, no dominant group, respect for minorities Freedom of the press
GCS	Newspapers can publish stories of their choosing without fear of censorship or retaliation When deciding upon policies and contracts, Government officials favor well-connected firms Effectiveness of national Parliament/Congress as a law making and oversight institution
HUM	Travel: domestic and foreign travel restrictions Freedom of political participation Imprisonments: Are there any imprisoned people because of their ethnicity, race, or their political, religious beliefs? Government censorship
PRS	Military in politics: The military are not elected by anyone, so their participation in government, either direct or indirect, reduces accountability and therefore represents a risk. The threat of military intervention might lead as well to an anticipated potentially inefficient change in policy or even in government. Democratic accountability: Quantifies how responsive government is to its people, on the basis that the less response there is the more likely it is that the government will fall, peacefully or violently. It includes not only if free and fair elections are in place, but also how likely it is that the government will remain in power.
RSF	Press freedom index
WMO	Institutional permanence: An assessment of how mature and well-established the political system is. It is also an assessment of how far political opposition operates within the system or attempts to undermine it from outside. Representativeness: How well the population and organized interests can make their voices heard in the political system.

Non-representative Sources	
AFR	Elections are free and fair
BTI	Stateness Political participation Institutional stability Political and social integration
CCR	Civil liberties Accountability and public voice
FHT	Political process: Deals with elections, referenda, party configuration, conditions for political competition, and popular participation in elections. Civil Society: Highlights the degree to which volunteerism, trade unionism, and professional associations exist, and whether civic organizations are influential. Independent Media: Press freedom, public access to a variety of information sources, and the independence of those sources from undue government or other influences.
GAL	Fairness of elections Satisfaction with democracy
LOB	Satisfaction with democracy Trust in parliament
MSI	Media sustainability index
USD	Trust in legislation Satisfaction with democracy
WCY	Transparency of government policy

Table B. Indicators for Political Stability

Political Stability	
Representative Sources	
DRI	Military coup risk: A military coup d'etat (or a series of such events) that reduces the GDP growth rate by 2% during any 12-month period. Major insurgency/rebellion: An increase in scope or intensity of one or more insurgencies/rebellions that reduces the GDP growth rate by 3% during any 12-month period. Political terrorism: An increase in scope or intensity of terrorism that reduces the GDP growth rate by 1% during any 12-month period. Political assassination: A political assassination (or a series of such events) that reduces the GDP growth rate by 1% during any 12-month period. Civil war: An increase in scope or intensity of one or more civil wars that reduces the GDP growth rate by 4% during any 12-month period. Major urban riot: An increase in scope, intensity, or frequency of rioting that reduces the GDP growth rate by 1% during any 12-month period.
EIU	Armed conflict Violent demonstrations Social unrest International tensions
GCS	Country terrorist threat: Does the threat of terrorism in the country impose significant costs on firms?
HUM	Frequency of political killings Frequency of disappearances Frequency of torture
IJT	Security Risk Rating

MIG	Extremism. The term “extremism” covers the threat posed by any individuals or organizations who hold a narrow set of fanatical beliefs. Extremists are likely to believe that any and all means are justified to eradicate the target of hostility, and are not afraid to destroy themselves in the process. This ideological aspect of extremism makes it highly unpredictable, and its close association with violence makes it highly dangerous. The extent to which extremism should be judged a threat to a particular business in a particular market can be assessed along the following lines: integration issues; religious tensions; pressure groups; terrorist activity; xenophobia.
PRS	Internal conflict: Assesses political violence and its influence on governance. External conflict: The external conflict measure is an assessment both of the risk to the incumbent government and to inward investment. Government stability: Measures the government’s ability to carry out its declared programs and its ability to stay in office. Ethnic tensions: This component measures the degree of tension within a country attributable to racial, nationality, or language divisions.
PTS	Political terror scale.
WMO	Civil unrest: How widespread political unrest is, and how great a threat it poses to investors. Demonstrations in themselves may not be cause for concern, but they will cause major disruption if they escalate into severe violence. At the extreme this factor would amount to civil war. Terrorism: Whether the country suffers from a sustained terrorist threat, and from how many sources. The degree of localization of the threat is assessed, and whether the active groups are likely to target or affect business.
Non-Representative Sources	
BRI	Fractionalization of political spectrum and the power of these factions. Fractionalization by language, ethnic and/or religious groups and the power of these factions. Restrictive (coercive) measures required to retain power. Organization and strength of forces for a radical government. Societal conflict involving demonstrations, strikes, and street violence. Instability as perceived by non-constitutional changes, assassinations, and guerilla wars.
WCY	Risk of political instability

Table C. Indicators for Government Effectiveness

Government Effectiveness	
Representative Sources	
DRI	Government instability: An increase in government personnel turnover rate at senior levels that reduces the GDP growth rate by 2% during any 12-month period. Government ineffectiveness: A decline in government personnel quality at any level that reduces the GDP growth rate by 1% during any 12-month period. Institutional failure: A deterioration of government capacity to cope with national problems as a result of institutional rigidity that reduces the GDP growth rate by 1% during any 12-month period.

EGV	Global E-government
EIU	Quality of bureaucracy Excessive bureaucracy/ red tape
GCS	Public spending composition Quality of general infrastructure Quality of public schools Time spent by senior management dealing with government officials
MIG	Quality of bureaucracy
PRS	Bureaucratic quality: Measures institutional strength and quality of the civil service, assess how much strength and expertise bureaucrats have and how able they are to manage political alternations without drastic interruptions in government services, or policy changes.
WMO	Policy consistency and forward planning: How confident businesses can be of the continuity of economic policy stance – whether a change of government will entail major policy disruption, and whether the current government has pursued a coherent strategy. Bureaucracy: An assessment of the quality of the country’s bureaucracy. The better the bureaucracy the quicker decisions are made and the more easily foreign investors can go about their business.
Non-representative Sources	
ADB	Management of public debt Policies to improve efficiency of public sector Revenue mobilization Budget management
AFR	Based on your experiences, how easy or difficult is it to obtain household services (like electricity or telephone)? Based on your experiences, how easy or difficult is it to obtain an identity document (like a birth certificate, passport)? Government handling of health services Government handling of education
ASD	Civil service Revenue mobilization and budget management Management and efficiency of public expenditures
BPS	How problematic are telecommunications for the growth of your business? How problematic is electricity for the growth of your business? How problematic is transportation for the growth of your business?
BRI	Bureaucratic delays
BTI	Consensus building Governance capability Effective use of resources
CPIA	Management of external debt Quality of public administration Revenue mobilization Budget management
FHT	Government and administration: Government decentralization, independent and responsibilities [sic] or local and regional governments, and legislative and executive transparency are discussed.
LBO	Trust in government
WCY	Government economic policies do not adapt quickly to changes in the economy The public service is not independent from political interference Government decisions are not effectively implemented Bureaucracy hinders business activity The distribution infrastructure of goods and services is generally

	inefficient Policy direction is not consistent
--	---

Table D. Indicators for Regulatory Quality

Regulatory Quality	
Representative Sources	
DRI	<p>Regulations – exports: A 2% reduction in export volume as a result of a worsening in export regulations or restrictions (such as export limits) during any 12-month period, with respect to the level at the time of the assessment.</p> <p>Regulations – imports: A 2% reduction in import volume as a result of a worsening in import regulations or restrictions (such as import quotas) during any 12-month period, with respect to the level at the time of the assessment.</p> <p>Regulations – other business: An increase in other regulatory burdens, with respect to the level at the time of the assessment, that reduces total aggregate investment in real LCU terms by 10%.</p> <p>Ownership of business by non-residents: A 1-point increase on a scale from “0” to “10” in legal restrictions on ownership of business by non-residents during any 12-month period.</p> <p>Ownership of equities by non-residents: A 1-point increase on a scale from “0” to “10” in legal restrictions on ownership of equities by non-residents during any 12-month period.</p>
EIU	<p>Unfair competitive practices</p> <p>Price controls</p> <p>Discriminatory tariffs</p> <p>Excessive protections</p>
GCS	<p>Administrative regulations are burdensome</p> <p>Tax system is distortionary</p> <p>Import barriers as obstacle to growth</p> <p>Competition in local market is limited</p> <p>Anti-monopoly policy is lax and ineffective</p> <p>Clusters are frequent</p> <p>Environmental regulations hurt competitiveness</p> <p>Complexity of tax system</p>
HER	<p>Regulation</p> <p>Government intervention</p> <p>Wage/prices</p> <p>Trade</p> <p>Foreign investment</p> <p>Banking</p>
MIG	<p>Unfair Competition.</p> <p>Unfair Trade.</p>
PRS	Investment profile
WMO	<p>Tax effectiveness: How efficient the country’s tax collection system is.</p> <p>Legislation: An assessment of whether the necessary business laws are in place.</p>
Non-representative Sources	
ADB	<p>Trade policy</p> <p>Competitive environment</p>

	Labor market policies
ASD	Trade policy and forex regime Enabling environment for private sector development
BPS	Information on the laws and regulations is easy to obtain How problematic are anti-competitive practices for the growth of your business? How problematic are unpredictable regulations for the growth of your business? How problematic are labor regulations for the growth of your business? How problematic are tax regulations for the growth of your business? How problematic are custom and trade regulations for the growth of your business?
BTI	Competition Price stability
FPIA	Competitive environment Trade policy
EBRD	Price liberalization Trade and foreign exchange system Competition policy
WCY	Access to capital markets (foreign and domestic) is easily available Ease of doing business Banking regulation does not hinder competitiveness Competition legislation in your country does not prevent unfair competition Customs' authorities do not facilitate the efficient transit of goods Financial institutions' transparency is not widely developed in your country Easy to start company Foreign investors are free to acquire control in domestic companies Price controls affect pricing of products in most industries Public sector contracts are sufficiently open to foreign bidders Real corporate taxes are non-distortionary Real personal taxes are non-distortionary The exchange rate policy of your country hinders the competitiveness of enterprises The legal framework is detrimental to your country's competitiveness Protectionism in your country negatively affects the conduct of business in your country Labor regulations hinder business activities Subsidies impair economic development

Table E. Indicators for the Rule of Law

Rule of Law	
Representative Sources	
DRI	Losses and costs of crime: A 1-point increase on a scale from "0" to "10" in crime during any 12-month period. Kidnapping of foreigners: An increase in scope, intensity, or frequency of kidnapping of foreigners that reduces the GDP growth rate by 1% during any 12-month period. Enforceability of government contracts: A 1 point decline on a scale from "0" to "10" in the enforceability of contracts during any 12-month period. Enforceability of private contracts: A 1-point decline on a scale from "0"

	to “10” in the legal enforceability of contracts during any 12-month period.
EIU	Violent crime Organized crime Fairness of judicial process Enforceability of contracts Speediness of judicial process Confiscation/expropriation
GCS	Common crime imposes costs on business Organized crime imposes costs on business Money laundering through banks is pervasive Quality of police The judiciary is independent from political influences of members of government, citizens or firms Legal framework to challenge the legality of government actions is inefficient Intellectual property protection is weak Protection of financial assets is weak Tax evasion
HER	Black market Property rights
HUM	Independence of judiciary
MIG	Organized crime Legal safeguards
PRS	Law and order. The law sub-component is an assessment of the strength and impartiality of the legal system, while the order sub-component is an assessment of popular observance of the law (assessed separately).
QLM	Direct financial fraud, money laundering and organized crime
WMO	Judicial independence: An assessment of how far the state and other outside actors can influence and distort the legal system. This will determine the level of legal impartiality investors can expect. Crime: How much of a threat businesses face from crime such as kidnapping, extortion, street violence and burglary.
Non-representative Sources	
ADB	Property rights
AFR	Based on your experiences, how easy or difficult is it to obtain help from the police when you need it?
ASD	Rule of law
BPS	Fairness, honesty, enforceability, quickness and affordability of the court system Property rights protection How problematic is organized crime for the growth of your business How problematic is judiciary for the growth of your business How problematic is street crime for the growth of your business
BRI	Enforceability of contracts
BTI	Rule of law Private property
CCR	Rule of law
CPIA	Property rights
FHT	Rule of law: Considers judicial/constitutional matters as well as the legal and de facto status of ethnic minorities.
GAL	Trust in the legal system
LBO	Trust in judiciary Trust in police

	Have you been a victim of crime?
WCY	Tax evasion is a common practice in your country Justice is not fairly administered in society Personal security and private property are not adequately protected Parallel economy impairs economic development in your country Patent and copyright protection is not adequately enforced in your country

Table F. Indicators for the Control of Corruption

Control of Corruption	
Representative Sources	
DRI	Risk event outcome non-price: Losses and costs of corruption: A 1-point increase on a scale from “0” to “10” in corruption during any 12-month period.
EIU	Corruption
GCS	Public trust in financial honesty of politicians Diversion of public funds due to corruption is common Frequent for firms to make extra payments connected to: import/export permits Frequent for firms to make extra payments connected to public utilities Frequent for firms to make extra payments connected to tax payments Frequent for firms to make extra payments connected to awarding of public contracts Frequent for firms to make extra payments connected to getting favorable judicial decisions Extent to which firms’ illegal payments to influence government policies impose costs on other firms
MIG	Corruption: There is an immense variety of activities that may be construed as corrupt. Bribery is the most obvious. However, what is and is not a bribe is a matter of presentation and perception in much the same way as “corruption” itself. Some of the issues that executives should consider include: accounting standards; anti-corruption policy credibility and enforceability; cronyism, nepotism and vested interests; cultural differences; judicial independence; transparency of decision-making.
PRS	Corruption. Measures corruption within the political system, which distorts the economic and financial environment, reduces the efficiency of government and business by enabling people to assume positions of power through patronage rather than ability, and introduces an inherent instability in the political system.
QLM	Indirect diversion of funds
WMO	Corruption: This index assesses the intrusiveness of the country’s bureaucracy. The amount of red tape likely to be encountered is assessed, as is the likelihood of encountering corrupt officials and other groups.
Non-representative Sources	
ADB	Transparency/corruption
AFR	How many elected leaders (parliamentarians or local councilors) do you think are involved in corruption? How many judges and magistrates do you think are involved in corruption? How many government officials do you think are involved in corruption? How many border/tax officials do you think are involved in corruption?

ASD	Anti-corruption
BPS	How common is it for firms to have to pay irregular additional payments to get things done? On average, what percent of total annual sales do firms pay in unofficial payments to public officials? How often do firms make payments to influence the content of new legislation? Extent to which firms' payments to public officials to affect legislation imposes costs on other firms? How problematic is corruption for the growth of your business? Frequency of bribery in utility, permits, procurement, health, fire inspection, environment, taxes, customs and judiciary
BRI	Internal causes of political risk: Mentality, including xenophobia, nationalism, corruption, nepotism, willingness to compromise, etc.
CCR	Transparency/corruption
CPIA	Transparency/corruption
FHT	Corruption
GAL	Frequency of corruption Frequency of household bribery
LBO	Have you heard of acts of corruption? Percentage of corrupt public officials
PRC	Corruption index
WCY	Bribing and corruption exist in the economy

Table G. Sources of Data

This table is taken from Table 1 of (Kaufmann et al., 2006a).

Abbreviation	Dataset	Source
ADB	Country Policy and Institutional Assessment	African Development Bank
AFR	Afrobarometer Survey	Afrobarometer
ASD	Country Policy and Institutional Assessments	Asian Development Bank
BTI	Bertelsmann Transformation Index	Bertelsmann Foundation
BPS	Business Enterprise Environment Survey	World Bank
BRI, QLM	Business Risk Service (BRI) Qualitative Risk Measure (QLM)	Business Environment Risk Intelligence
CDU	State Capacity Project	Columbia University
CPIA	Country Policy & Institutional Assessment	World Bank
EIU	Country Risk Service	Economist Intelligence Unit
EBR	Transition Report	European Bank for Reconstruction and Development
FRH, FHT, CCR	Freedom in the World (FRH) Freedom House Nations in Transit (FHT) Countries at the Crossroads (CCR)	Freedom House
GAL	The Voice of the People Survey	Gallup International
EGV	Global E-Governance Index	Brown University's Center for Public Policy

DRI, WMO	Global Risk Service (DRI) Business Conditions and Risk (WMO)	Global Insight
HER	Economic Freedom Index	Heritage Foundation
IJT	Country Security Risk Assessment	IJET Travel Intelligence
WCY	World Competitiveness Yearbook	Institute for Management and Development
MSI	Media Sustainability Index	International Research and Exchanges Board
LBO	Latinobarometro Surveys	Latinobarometro
MIG	Grey Area Dynamics	Merchant International Group
PRC	Corruption in Asia	Political Economic Risk Consultancy
PRS	International Country Risk Guide	Political Risk Services
RSF	Reporters Without Borders	Reporters Without Borders
TPR	Trafficking in People Report	State Department
HUM	Human Rights Dataset	State Depart/ Amnesty International
WBS	World Business Environment Survey	World Bank
PIA	Country Policy and Institutional Assessment	World Bank
GCS	Global Competitiveness Report	World Bank

References

- Adcock, R. and D. Collier (2001) 'Measurement Validity: A Shared Standard for Qualitative and Quantitative Research', *The American Political Science Review* 95(3): 529-546.
- Andres, A. R. (2006) 'Software Piracy and Income Inequality', *Applied Economics Letters* 13(2): 101-105.
- Apodaca, C. (2004) 'The Rule of Law and Human Rights', *Judicature* 87(6): 292-299.
- Arndt, C. and C. Oman (2006) 'Uses and Abuses of Governance Indicators'. OECD Development Centre.
- Bartholomew, D. J. (1995) 'Spearman and the Origin and Development of Test Theory', *British Journal of Mathematical and Statistical Psychology* 48: 211-220.
- Bollen, K. A. (1989) *Structural Equations with Latent Variables*. New York: Wiley.
- Bollen, K. A. (1993) 'Liberal Democracy: Validity and Method Factors in Cross-National Measures', *American Journal of Political Science* 37(4): 1207-1230.
- Burnside, C. and D. Dollar (2004) 'Aid, Policies, and Growth', *American Economic Review* 94(3): 781-784.
- Campbell, D. T. and D. W. Fiske (1959) 'Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix', *Psychological Bulletin* 56: 81-105.
- Carmines, E. G. and R. A. Zeller (1979) *Reliability and Validity Assessment*. Beverly Hills: Sage Publications.
- Clarke, G. R. G. and S. J. Wallsten (2006) 'Has the Internet Increased Trade? Developed and Developing Country Evidence', *Economic Inquiry* 44(3): 465-484.
- Cronbach, L. and P. Meehl (1955) 'Construct Validity in Psychological Tests', *Psychological Bulletin* 52(4): 281-302.
- Das, G. G. and S. Andriamananjara (2006) 'Hub-and-spokes Free Trade Agreements in the Presence of Technology Spillovers: An Application to the Western Hemisphere', *Review of World Economics* 142(1): 33-66.
- Dunn, J. M., R. Schneck and J. Lawson (1973) 'A Test of the Uni-Dimensionality of Various Political Scales through Factor Analysis: A Research Note', *Canadian Journal of Political Science* 6(4): 664-669.
- Easterly, W., R. Levine and D. Roodman (2004) 'Aid, Policies and Growth: Comment', *American Economic Review* 94(3): 774-780.
- Eid, M., T. Lischetzke, F. W. Nussbeck and L. I. Trierweiler (2003) 'Separating Trait Effects from Trait-Specific Method Effects in Multitrait-Multimethod Models: A Multiple-Indicator CT-C(M-1) mModel', *Psychological Methods* 8: 36-60.
- Elkins, Z. (2000) 'Gradations of Democracy? Empirical Tests of Alternative Conceptualizations', *American Journal of Political Science* 44(2): 293-300.
- Faber, J. (1987) 'Measuring Cooperation, Conflict and Social Network of Nations', *The Journal of Conflict Resolution* 31(3): 438-464.
- Fallon, J., Richard H. (1997) "'The Rule of Law" as a Concept in Constitutional Discourse', *Columbia Law Review* 97(1): 1-56.
- Forsythe, G. B., W. C. McGaghie and C. P. P. Friedman (1986) 'Construct Validity of Medical Clinical Competence Measures: A Multitrait-Multimethod Matrix Study

- Using Confirmatory Factor Analysis', *American Educational Research Journal* 23(2): 315-336.
- Hammond, K. R., R. M. Hamm and J. Grassia (1986) 'Generalizing Over Conditions by Combining the Multitrait-Multimethod Matrix and the Representative Design of Experiments', *Psychological Bulletin* 100: 257-269.
- Hart, D., R. Atkins and J. Youniss (2005) 'Knowledge, Youth Bulges, and Rebellion', *Psychological Science* 16(8): 661-662.
- Hattie, J. (1984) 'An Empirical Study of Various Indices for Determining Unidimensionality', *Multivariate behavioral Research* 19: 49-78.
- Hirschman, A. O. (1970) *Exit, Voice, and Loyalty*. Cambridge, MA: Harvard University Press.
- Jung, M. (2006) 'Host Country Attractiveness for CDM Non-sink Projects', *Energy Policy* 34(15): 2173-2184.
- Kaufmann, D., A. Kraay and M. Mastruzzi (2004) 'Governance Matters III: Governance Indicators for 1996, 1998, 2000, and 2002', *The World Bank Economic Review* 18(2): 253-288.
- Kaufmann, D., A. Kraay and M. Mastruzzi (2005) 'Governance Matters IV: Governance Indicators for 1996-2004'. Washington, D.C.: World Bank.
- Kaufmann, D., A. Kraay and M. Mastruzzi (2006a) 'Governance Matters V: Aggregate and Individual Governance Indicators for 1996-2005'. Washington, D.C.: World Bank.
- Kaufmann, D., A. Kraay and M. Mastruzzi (2006b) 'The Worldwide Governance Indicators Project: Answering the Critics'. Washington, D.C.: World Bank.
- Kaufmann, D., A. Kraay and Zoido-Lobaton (1999a) 'Aggregating Governance Indicators'. Washington, D.C.: World Bank.
- Kaufmann, D., A. Kraay and Zoido-Lobaton (2002) 'Governance Matters II: Updated Indicators for 2000/01'. Washington, D.C.: World Bank.
- Kaufmann, D., A. Kraay and P. Zoido-Lobaton (1999b) 'Governance Matters'. Washington, D.C.: World Bank.
- Knack, S. (2006) 'Measuring Corruption in Eastern Europe and Central Asia: A Critique of the Cross-Country Indicators'. Washington, D.C.: World Bank.
- Laughland, A. S., W. N. Musser, J. S. Shortle and L. M. Musser (1996) 'Construct Validity of Averting Cost Measures of Environmental Benefits', *Land Economics* 72(1): 100-112.
- Liu, M. C. and G. San (2006) 'Social Learning and Digital Divides: A Case Study of Internet Technology Diffusion', *Kyklos* 59(2): 307-321.
- Llamazares, I. (2005) 'Patterns in Contingencies: The Interlocking of Formal and Informal Political Institutions in Contemporary Argentina', *Social Forces* 83(4): 1671-1695.
- Millennium Challenge Account (2006). *Report on the Criteria and Methodology for Determining the Eligibility of Candidate Countries for Millennium Challenge Account Assistance in FY 2006* 2006 [cited August 26, 2006 2006]. Available from http://www.mca.gov/about_us/congressional_report/FY06_Criteria_Methodology.pdf.

- Nandakuma, R. and W. Stout (1993) 'Refinement of Stout's Procedure for Assessing Latent Trait Unidimensionality', *Journal of Educational Statistics* 18(1): 41-68.
- Neumayer, E. (2002) 'Do Democracies Exhibit Stronger International Environmental Commitment? A Cross-Country Analysis', *Journal of Peace Research* 39(2): 139-164.
- North, D. and R. P. T. Thomas (1996) *Rise of the Western World: A New Economic History*. Cambridge: Cambridge University Press.
- Patterson, M. L. (1990) 'On the Construct Validity and Developmental Course of Rapport', *Psychological Inquiry* 1(4): 320-321.
- Perry, J. L. (1996) 'Measuring Public Service Motivation: An Assessment of Construct Reliability and Validity', *Journal of Public Administration Research and Theory: J-PART* 6(1): 5-22.
- Segars, A. (1997) 'Assessing the Unidimensionality of Measurement: a Paradigm and Illustration Within the Context of Information Systems Research', *Omega, International Journal of Management Science* 25(1): 107-121.
- Smith, G. T. (2005) 'On Construct Validity: Issues of Method and Measurement', *Psychological Assessment* 17(4): 396-408.
- Studemund, A. H. (1997) *Using Econometrics: A Practical Guide*. 3rd edition ed: Addison-Wesley Educational Publishers.
- Wallace, R. B. and A. R. Herzog (1995) 'Overview of Health Measures in the Health and Retirement Study', *The Journal of Human Resources* 30(S84-S107).
- Westen, D. and R. Rosenthal (2003) 'Quantifying Construct Validity: Two Simple Measures', *Journal of Personality and Social Psychology* 94: 608-618.
- White, H. (1982) 'Maximum Likelihood Estimation of Misspecified Models', *Econometrica* 50(1): 1-26.
- Williams, R. H., D. W. Zimmerman, B. D. Zumbo and D. Ross (2003) 'Charles Spearman: British Behavioral Scientist', *Human Nature Review* 3: 114-118.
- Wolfensohn, J. (1999) 'Remarks at a Global Forum on Fighting Corruption'. Washington, D.C.: World Bank.
- Wooldridge, J. M. (2003) *Introductory Economics: A Modern Approach*. N.p.: South-Western.
- World Bank (1994) 'Governance: The World Bank's Experience'. Washington, D.C.: World Bank.
- World Bank (1998) *Assessing Aid -- What Works, What Doesn't, and Why*. Washington, D.C.: World Bank.
- World Bank (2006a) 'Strengthening Bank Group Engagement on Governance and Anticorruption'. Washington, D.C.: World Bank.
- World Bank (2006b) 'World Bank Releases Largest Available Governance Data Source'. Washington, D.C.: World Bank.