United Nations University

World Governance Survey Discussion Paper 2
August 2002

ASSESSING GOVERNANCE: METHODOLOGICAL CHALLENGES

Julius Court, Goran Hyden and Ken Mease[1]

Introduction

The first World Governance Survey (WGS) Discussion Paper[2] outlined the increasing importance attributed to governance in the international community. If governance matters, so does the need for more reliable and valid data. Yet, many challenges remain in effectively assessing and analyzing governance issues. There remains debate over how best we can meaningfully measure governance – many believe current indicators provide poor measures of key governance processes. Similarly, there remains a lack of agreement over who is best placed to provide insights as to the quality of governance in a particular country and how it compares to the situation in other countries. These are critical issues. Without advance on such concerns, it will not be possible to assess how governance varies across the world, what role governance really plays in development, and what aspects of it may be particularly critical. Based on our experiences of collecting data in 22 countries, this paper is intended to spur discussion on ways to address these challenges.

We begin this paper by assessing different ways in which governance has been studied by others so as to place our own study in a comparative methodological context. We proceed by discussing the implementation of our own survey, highlighting some of the problems we also

encountered. The next section deals with the way we have analyzed our data. In conclusion, we make an overall assessment of the value of the data that we present and analyze in subsequent papers.

Assessing Governance Empirically

Measuring issues of governance poses challenges that are not encountered in the economic or social development fields. While it is easier to provide firm indicators of such things as economic growth or primary school enrolment, it is much more difficult to find and agree upon indicators of a political macro phenomenon like governance or political rights. Perhaps because it is a broad and complicated concept, there exists no regular, systematic and cohesive data collection effort centered on the concept of governance.

In short, there is very little objective data for many countries – and even less that is comparative in a meaningful way. Most existing sources of governance data are subjective (see appendix).[3] Kaufmann et al.[4] highlight a number of reasons why it is useful to gather data on governance perceptions although the data collected is inherently subjective. For example, perceptions may often be more meaningful than objective data, especially when it comes measuring the public faith in institutions.

Trying to overcome these shortcomings or challenges is not easy, but because there is a growing interest in trying to link perceptions of governance with development outcomes, we believe that it is important to address them and see how far more meaningful comparative data can be generated. Other organizations, such as the Economist Intelligence Unit, Political Risk Services and Freedom House, that are collecting this type of data already, use a small number of country, sector, and regional experts to generate consensus ratings. Some others undertake cross-country surveys of citizens, e.g. Gallup International, or the business

---

[2] Goran Hyden and Julius Court, 2002, *Governance and Development, World Governance Survey Discussion Paper 1*, Tokyo: UNU.
[3] While more objective data would be helpful, it is unclear how such data – especially on governance processes – could be generated. It is certainly worthwhile trying to analyze and measure such qualitative phenomenon in a scientific, empirical manner. And even though such data should be treated with care, they can providing interesting insights and stimulate valuable discussion. In addition, subjective data is also important since perceptions of efficiency and legitimacy do matter.
[4] D. Kaufmann, A. Kraay, and P. Zoido-Lobaton. "Aggregating Governance Indicators", *Policy Research Paper No 2195*. Washington D.C.: The World Bank 1999.

community, such as World Economic Forum. Kaufmann, Kraay and Zoido-Lobaton provide the most useful aggregate scale for measuring governance by drawing on the most credible available sources. [5]

These different approaches have various methodological advantages and benefits.[6] Expert polls have the advantage of providing cross-country comparability, but suffer from being based on the opinions of only a few experts per country. There are at least two problems with this type of polls. One is that the ratings tend to reflect the political or ideological agenda of the organization providing the ratings. The second is that countries with good economic outcomes tend to receive an automatic high rating[7]. The advantage of surveys is that they reflect the opinions of a larger number of respondents that are more closely connected with the countries that they are assessing. Nonetheless, they also have their disadvantages. Survey questions tend to be interpreted in context-specific ways, thus limiting the cross-country comparability of responses. In addition, they tend to be costly to design and implement and, as a result, typically cover a much smaller set of countries than the polls of experts.

The Kaufmann et al. study identifies three specific methodological problems with existing data: (1) ill-defined or broad concepts and poorly worded questions; (2) available indicators provide poor measures for some key governance issues; and (3) the tendency to aggregate measures by combining indicators from different sources does not enhance precision.

While there have been significant advances in recent years, existing data suffer from a number of limitations in terms of providing a cohesive picture of governance – particularly for developing countries. For example, the Economist Intelligence Unit (EIU) and the Global Competitiveness Survey (GCS) are primarily concerned with indicators related to economic development, International Country Risk Guide (ICRG) foremost with issues of interest to business corporations and potential investors. Scholars and practitioners frequently use the Freedom House Index (FHI) and Polity datasets to measure the level of democracy in a given country, but these deal only with a specific set of civil freedoms and political rights. Transparency International (TI) reports on corruption in the private sector and state. There

[5] *Idem*.
[6] See: M. Blagescu, J. Court, G. Hyden, K. Mease, and K. Suzuki, "Assessing and Analysing Governance: Lessons from the World Governance Survey Pilot Phase", *WGS Working Paper No 2*, Tokyo: United Nations University, August 2001.

have been suggestions that the Transparency International work has methodological problems, especially with regard to data collection. Furthermore, They have a very narrow measure of governance.

Another shortcoming is that few sources of governance data cover the whole world. Most are confined to a limited sample of countries. Exceptions are the EIU, FHI, and ICRG, but, as suggested above, these do not provide comprehensive thematic coverage of governance issues. The third problem is that each institution uses a different methodology to collect data. Some use panels of external experts, while others conduct surveys. Those that provide global coverage tend to rely on panels of experts from outside the country.

Despite considerable efforts, what is still missing are "holistic" assessments of governance that take into account the key processes that are relevant. Thus, among others, Kaufmann and his colleagues highlight the need to improve and extend the cross-country survey work on governance perceptions, as well as to employ country-specific in-depth governance diagnostics.[8]

It is important to make a distinction between governance *performance* indicators and governance *process* indicators. Performance indicators refer to the quality of governance in terms of a normative outcome, such as the level of corruption. Process indicators refer to the quality of governance in terms of how outcomes are achieved. As highlighted in Discussion Paper One, an emphasis on governance process stems from a human rights approach to development. While there has been a considerable improvement in the monitoring of governance outcomes, particularly over the last 10 years, the monitoring of processes has remained much more limited. The challenge, therefore, is how to measure governance cohesively and systematically in terms of critical *processes*.

---

[7] Daniel Kaufmann, A. Kraay, and P. Zoido-Lobaton. "Governance Matters." *Policy Research Paper No 2196.* Washington D.C.: The World Bank 1999.
[8] *Idem*, p. 28.

Data Collection Strategies

With our goal to collect systematic data on a comprehensive range of governance issues, we had several options. We considered using an international panel of experts or well-informed persons. This would have allowed us to generate data for cross-country comparisons, but it would have been a replication of what others, like the Freedom House, is already doing. We also wanted to avoid the major shortcomings with such panels, namely their relative superficiality stemming from the coverage of a large number of countries within a limited time frame. We wanted to do something that others have not yet done: collect the opinions of people resident in the countries under consideration, while at the same time come up data for comparative purposes.

We also considered the possibility of conducting a public opinion survey. This approach would have been the most desirable in some ways, because it would no doubt have helped us generate credible data on public perceptions of governance. We had to reject this approach on three grounds. The first reason for rejecting this approach was simply that the public would not be able to give credible responses to many of the questions we were asking since they required a high degree of detailed knowledge and understanding of specific governance issues. We would not be able to generate the type of data we wanted from the public at large. The second reason for not embarking on a public opinion survey was the problem of drawing representative samples in many countries. The infrastructure for selecting such samples and how to determine what is representative of the population discouraged us from thinking about this approach. It would only be possible to undertake credible public opinion surveys in about sixty more developed countries. It would be impossible to collect reliable data in many of the less-developed countries that are, in many ways, the most relevant and interesting for our purposes. The final reason was simply the cost of conducting such surveys. According to Gallup International, such a survey would cost somewhere in the range of $50,000 per country. We could not justify such expenses, especially since it would not be possible to ask detailed governance questions to a cross-section of the public in each country.

We had plans to conduct focus group discussions in three countries as a way of learning what type of data that approach will generate. Unfortunately, in the end, we have a very limited experience to draw from since data were submitted only from two countries—the Philippines

and Barbados.[9] Nevertheless, a few things came out very clearly, especially from the Philippine study. In terms of strengths, it is clear that the approach can generate information about the background conditions that determine certain ratings. One obtains a much better sense of the independent variables that determine governance ratings. A second advantage is that, because it is highly participatory, it has the potential of generating solutions to the problems identified by the group members. The focus group approach, however, also has certain disadvantages that cannot be ignored. It is very demanding and requires very skilled coordinators. We are not sure that it would be possible to find such coordinators in many countries. Second, since the approach catalyzes collective integrated thinking, it makes individual ratings insignificant. Another drawback is that accuracy suffers, as some individuals may not feel comfortable to speak up in public. A third point is that, although it generates more location-specific data, the focus group approach yields less systematic results. For instance, in the Philippine study, there were marked differences between groups depending on social background and geographic location, e.g. Luzon versus Mindanao. While we do not reject the idea that the focus group approach may constitute a complementary approach to data collection, we decided, on balance, that it would not serve our purposes well. It simply would have left us with more questions than answers when it comes to assessing the data. Our assessment in this regard reflects the consensus view of most researchers, i.e. that focus groups are best used to identify issues and develop surveys rather than as the only source of data.

In the end, we opted an approach for data collection data by designing a survey aimed at interviewing a cross-section of well-informed persons (WIPs) in each country. These were individuals who are experienced in and informed about the governance realm. The premise of our approach was that they would be able to provide the most knowledgeable ratings about governance as well as qualitative comments to back up their assessments.

---

[9] For a comparison of ratings between the focus groups approach and WIP approach in the Philippines, please refer to Appendix 5 in Blagescu et al, 2001, *op cit*.

Implementation of the Survey

The survey was conducted in late 2000 and early 2001 in twenty-two countries that can be described as transitional societies[10]. A country coordinator was identified to implement the survey in each country. These coordinators were in some cases heads of local research and policy institutes working on governance and/or development issues, in others senior researchers located at local universities. The latter were mostly political scientists or economists. The country coordinators were paid between US$2,000 and US$3,000 to deliver thirty-five completed questionnaires and to prepare a report.[11]

The project instrument was administered in eight different languages. Translation was done either at the UNU headquarters or by the country coordinators themselves. Various modes were used to collect the data – face-to-face interviews, faxes, and emails. The first option was the most common, but in some countries, the others were also used. For the full questionnaire, see Appendix 3.

Our instructions to the country coordinators emphasized that they should select respondents randomly from a cross-section of persons representing different perspectives on governance. They should be at least 35 years of age and should have significant experience in public life. More specifically, we asked them for a rough balance between the following groups:

- High ranking civil servants
- Long-standing parliamentarians
- Business persons
- Senior judges and lawyers
- Respected academics, consultants or policy advisors
- Heads or senior officials in local NGOs
- Editors or senior reporters in the media
- Any other relevant category

---

[10] Those countries where (in alphabetical order): Argentina, Barbados, Bulgaria, Chile, China, India, Indonesia, Jordan, Kyrgyzstan, Mongolia, Nepal, Nigeria, Pakistan, Papua-New Guinea, Peru, Philippines, Russia, Samoa, South Korea, Tanzania, Thailand and Togo.
[11] In addition, they were required to transcribe the open-ended comments and enter the results of the completed surveys into an Excel spreadsheet.

It turned out that it was easier to find such a balance among the cross-section of well-informed persons in some countries than in others. Country coordinators also found out that the response to their request was not automatic. In our survey, the response rates varied from over 70 percent (China, Jordan, Kyrgyzstan, Samoa, Thailand, and Togo) to a low of 31 percent in Chile. The average response rate was 57 per cent. The details on each country are contained in Table 1.

Because of the unevenness of the data, we were in the end unable to use all of what we collected. We had to remove six 'poor data-quality' countries – Barbados, Korea, Nepal, Nigeria, Papua-New Guinea and Samoa. Four reasons were used to make this decision. The first was that completed questionnaires contained too many missing values. The second was too skewed a balance in the distribution of WIPs. Some countries had too many missing WIP groups, while others lacked an acceptable balance between respondents from state and society. We decided that WIPs associated with the state should make up at least 20 percent but not exceed 40 percent. Finally, in some countries, such as Barbados, the sample size was just too small. If these criteria were not met, a country was removed from the sample. This left us with 16 countries of sufficiently high data quality to be included here. The response rate for 14 of the 16 countries where we currently have response rate information was 57%.

Table 1. Response Rates and Language used for the WGS

	Country	Language	Sample Released	Completed (Sample Size)	Response Rate
Africa					
	Togo	French	54	42	0.78
	Tanzania	English	55	33	0.60
	Nigeria	English	60	38	0.63
Asia					
	China	Chinese	45.	37	0.82.
	India	English	90	38	0.42
	Indonesia	English	55	35	0.64
	Mongolia	Mongolian	N/A	41	
	Nepal	English	72	37	0.51
	Pakistan	English	110	37	0.34
	Philippines	English	70	35	0.50
	PNG	English	N/A	37	
	Samoa	English	50	37	0.74
	Thailand	Thai	50	42	0.84
Eastern Europe					
	Bulgaria	Bulgarian	106	42	0.40
	Russia	Russia	N/A	39	
	Kyrgyz Republic	Russia	48	40	0.83
Middle East					
	Jordan	Arabic	55	40	0.73
Latin America					
	Argentina	Spanish	106	38	0.36
	Chile	Spanish	113	35	0.31
	Barbados	English	46	23	0.48
	Peru	Spanish	80	37	0.46
OECD					
	Korea	Korean	60	41	0.68

WGS Indices

The WGS questionnaire is comprised of thirty indicators. The items are equally divided into six sections covering the six governance arenas mentioned above, i.e. with 5 questions per arena. Respondents were asked to rate various issues concerning governance using the same five-point response scale: as either very high, high, moderate, low, or very low. Table 2 presents the median, mean, standard deviation, and sample size for each item comprising the World Governance Survey (WGS2000) and the views of our respondents, subsequently labeled WGS1996 indices in this paper

Table 2. Medians, Means, and Standard Deviations for WGS2000, WGS1996 and the Six Governance Arenas

DESCRIPTION	Median	Mean	Std. Dev.	Sample Size
WGS in 1996	81	80.17	17.03	587
WGS in 2000	85	84.26	18.93	587
Civil Society in 1996	15	14.49	3.13	587
Civil Society in 2000	15	15.18	3.49	587
Political Society in1996	13	12.51	3.69	587
Political Society in 2000	13	13.22	4.19	587
Executive in 1996	14	14.14	4.06	587
Executive in 2000	15	14.57	4.27	587
Bureaucracy in 1996	13	12.93	3.57	587
Bureaucracy in 2000	13	13.22	3.84	587
Economic Society in 1996	14	13.63	3.33	587
Economic Society in 2000	15	14.55	3.44	587
Judiciary in1996	12	12.49	3.67	587
Judiciary in 2000	13	13.52	3.83	587

Reliability

Reliability is a fundamental issue in all research and is especially important in psychological measurement.[12] Scale or index reliability is the proportion of variance attributable to the true score of the latent variable, which in this case is governance. Reliability and statistical power are interrelated: as reliability increases, so does the statistical power of the scale. Reliability is

---

[12] Robert DeVellis, *Scale Development: Theory and Applications*, Newbury Park, Cal., Sage Publications, 1991.

inversely related to errors of measurement; the larger the error, the worse the reliability.[13] One way to increase reliability is to increase the number of scale items. In other words, scales with more items are likely to generate greater internal consistency. The most common statistical approach to measure reliability is Cronbach's *Alpha*.[14] Internal consistency is attained when the items, designed to measure the same construct, interrelate with one another.[15] The number of items in the scale or index affects Cronbach's *Alpha*. Generally, the more items in a scale or subscale, the higher the *Alpha* and, therefore, the higher the reliability.

The WGS indices consist of twenty-eight positively and two negatively worded items. The number of questions in the questionnaire is reasonable, with only a couple of WIPs complaining that it was too long. The two negatively worded items (#3 and #23), after reversing the values, were negatively correlated with every other item in the WGS. This may suggest that respondents engaged in what is referred to as "satisficing"[16].

Satisficing occurs when a respondent does not pay close attention to survey questions and takes cognitive shortcuts. Because the first negatively worded item appears in the third questions, it is doubtful that satisficing occurred regularly. Sometimes, respondents will answer in the same way to multiple items, or, in some cases, to the entire survey. An examination of the individual responses from each country uncovered only a couple of cases where a respondent entered the same value for every question in the survey. While there is usually some satisficing in all surveys, there is no indication that the WGS survey had an excessive amount. Country coordinators also regularly reported that they had been unsure about the analysis of these questions. After a careful systematic analysis of these two items we believe some respondents misinterpreted them, but the evidence does not suggest that all respondents misinterpreted these two negatively worded items. Therefore, we have decided to leave them in our study and include them in all calculations, rather than second-guess our respondents.

---

[13] Paul Spector, *Summated Rating Scale Construction: An Introduction*. Newbury Park, Cal., Sage Publications, 1992.
[14] Andrew Comrey, "Factor-Analytic Methods of Scale Development in Personality and Clinical Psychology", *Journal of Consulting and Clinical Psychology*, 56 (5), 1988, pp: 754-761.
[15] George Bohrnstedt, "A Quick Method of Determining the Reliability and Validity of Multiple-Item Scales", *American Sociological Review* 34 (4), 1969, pp: 542-548; also, Spector *op.cit.*
[16] Jon Krosnick, "Satisficing in Surveys: Initial Evidence" (pp, 29-44) in M.T. Braverman and J.K. Slater, eds., *New Directions in Evaluations*. San Francisco, Jossey-Bass, 1996.

Thirty items were used to calculate the 1996 and 2000 WGS governance scores. Both indices have a minimum value of 30 and a maximum value of 150. Simply put, the higher the number, the better the perception of governance. The results presented in Table 3 suggest that the WGS index in 1996 and 2000 exhibit very high levels of reliability. Both in 2000 and in 1996 Cronbach's A*lpha* score is a very high .93. (A perfect score is 1.00.) The WGS *Alpha* scores are well above the threshold of .70 suggested by Nunnally; however, scales with an *Alpha* score of .60 and above are commonly used and reported in major academic journals.[17] Removing items q3 and q23 raised the score only slightly.

We decided to examine the reliability of the individual governance arenas to more fully understand the impact of q3 and q23. In Table 3 we see that when we remove q3 and q23 in the civil society and economic society arenas the *Alphas* scores shoot up dramatically and are above Nunnally's .70 threshold. With only 5 items in each arena, the effect of a negatively correlated item is very strong. In conclusion, while the arenas with q3 and q23 are somewhat below Nunnally's cut off, the overall WGS indices are demonstrating exceptionally high reliability.

---

[17] Jamel C. Nunnally, *Psychometric Theory*, New York: McGraw Hill, 1978

Table 3. Cronbach's Alpha Reliability Scores for WGS 2000 and WGS 1996

WGS Indices	Cronbach's *Alpha*	Cronbach's *Alpha* w/o q3 and q23
WGS in 1996	93	.93
WGS in 2000	93	94
Civil Society in 1996	*.52*	*71*
Civil Society in 2000	*.64*	*.72*
Political Society in 1996	.77	N/A
Political Society in 2000	.80	N/A
Executive in 1996	.77	N/A
Executive in 2000	.79	N/A
Bureaucracy in 1996	.75	N/A
Bureaucracy in 2000	.77	N/A
Economic Society in 1996	*.67*	*.74*
Economic Society in 2000	*.69*	*.76*
Judiciary in 1996	.77	N/A
Judiciary in 2000	.77	N/A

Validity

A scale or index is valid if it measures what it was designed to measure. Validation usually involves testing a hypothesis about the scale or index. The test that best fits the WGS is the "known groups" validity test. In the case of the WGS we might hypothesize that countries that are considered to have high levels of governance will score higher on the WGS than countries considered to have lower levels. Similarly, the same may be true of different groups of WIPs groups.

Test One - World Bank Governance Indicators. In Table 4, the WGS2000 and WGS1996 are compared to governance indicators developed by Kaufmann and his colleagues at the World Bank.[18] While some of the countries, such as Togo, match up quite well in this comparison, others, such as Indonesia and the Philippines, do not. In these specific cases, the difference is

due to events that transpired after the Kaufmann collected his data and before the WGA data was collected.

Table 4. Median WGS 2001, WGS 1996, World Bank Governance Index and the Kaufmann et. al. Governance Ranking for the 16 Countries[19]

Country	Sample Size	WGS in 1996	WGS in 2001	World Bank Governance Index 2000	Kaufman Governance Ranking 2000
Togo	42	67.5	62.5	-4.2	136
Pakistan	33	68.0	65.0	-3.6	133
Russia	38	72.0	73.0	-3.2	129
Kyrgyzstan	39	83.0	75.0	-2.5	118
Philippines	**35**	**89.0**	**75.0**	**1.3**	**63**
Indonesia	**35**	**64.0**	**80.0**	**-4.6**	**143**
China	33	71.0	82.0	-1.2	94
Peru	37	65.0	82.0	-1.1	91
Argentina	35	80.0	83.0	2.0	54
Bulgaria	41	76.5	83.0	0.0	75
Mongolia	39	86.0	86.0	1.3	62
Tanzania	33	85.5	91.0	-0.8	87
Jordan	40	93.0	97.5	2.0	53
India	36	96.5	98.0	0.0	76
Chile	30	94.0	99.0	5.3	28
Thailand	41	90.0	100.0	0.9	65

The results of the correlation test on the 2000 WGS scores and the World Bank's Governance index was .72 with Indonesia and the Philippines included in the analysis, and .84 with them removed. In the Philippines the WGA scores reflect a downturn in perceptions of governance caused by the Estrada regime's well-publicized problems. In Indonesia, the higher WGS scores are the result of the end of the Suharto regime.

Test Two - State WIPs versus Non-State WIPs. We hypothesized that government actors would have higher evaluations of governance than WIPs outside state institutions. In Table 6 we can see that WIPs working in State institutions (high ranking government officials, members of parliament, and high ranking civil servants) do indeed have significant higher governance scores than other WIPs. Clearly, the results of the correlation analysis of the

[18] D. Kaufmann et al "Aggregating Governance Indicators", *op.cit*.

World Bank governance scores and the ANOVA results of State WIP groups suggest that the WGA passes the 'known groups' validity test.

Analyzing the Responses Across Expert Group

Across the countries the scores of certain experts were significantly higher than other groups. Table 5 reports the median WGA2001 and WGA1996 scores for eleven expert groups. Using the information from individual questionnaires, these are broken down further than the broader classifications indicated in the instructions to country coordinators.

Table 5. Median Values for WGS2000 and WGS1996 of WIP Groups from all 16 Countries

Category of WIP	WGS2000	WGS1996	Sample Size
Government	95.0	84.0	71
Parliament	90.0	83.5	58
Civil Service	88.0	88.0	41
Academics	85.0	80.0	71
International Org.	84.5	79.0	30
Legal	82.5	83.0	56
NGOs	81.0	78.5	69
Business	80.0	77.0	64
Religious*	79.0	82.0	9
Media	80.0	82.0	37
Other	80.0	79.0	81

* very small sample size

The results presented in Table 6 suggest that WIPs in government, the parliament and the civil service, rate the state of governance significantly higher than other WIP groups. Other groups, such as WIPs in NGOs, business, international organizations, the media, and academia gave significantly lower scores, which offset these high scoring groups. We

[19] This table is arranged with the lowest scoring country, Togo, at the top.

expected that WIPs directly connected to the state would likely have higher perceptions of governance. One can speculate that the three high scoring groups may suffer from self-evaluation bias. Another way to view these results is that WIPs in international organizations, business, the media, and academia are usually more critical of the state of governance. This is an area that provides basis for interesting and insightful analysis of perceptions of governance among different groups of WIPs.

Table 6. ANOVA - Differences in the WGS2000 Scores between WIP Groups

WIP Groups	Difference in Mean WGS 2000 Score	Significance
Government higher than Academics	10.9	0.05
Government higher than Business	12.7	0.05
Government higher than NGO	14.0	0.05
Government higher than International Org	13.9	0.05
Government higher than Legal	9.9	0.05
Government higher than Media	12.7	0.05
Parliament higher than Legal	7.0	0.05
Parliament higher than Media	9.7	0.05
Parliament higher than Academics	8.0	0.05
Parliament higher than Business	9.8	0.05
Parliament higher than NGO	11.1	0.05
Parliament higher than International Org	10.9	0.05
Civil Service higher than Business	7.9	0.05
Civil Service higher than NGO	9.2	0.05
Civil Service higher than International Org	9.2	0.05

We also used ANOVA to examine the potential differences between experts within each country and found that only three countries had significant difference – Togo, Tanzania, and Mongolia – at the .05 level. Overall, we believe that the mix of state and non-state WIP groups provide a balanced perception of governance consistent with the WGA2001's high correlation with the World Bank governance indicators discussed above.

Comparison with Other Measures of Governance

We have now reached the point where we want to report on how the WGS findings compare to other existing measures of governance. In this section, we compare the WGS with data from the World Bank, Freedom House Index, International Country Risk Guide (ICRG), and Polity IV. It is important to reiterate that our governance assessment is based on the subjective perceptions of local well-informed persons. Most other measures use panels of international experts, businesspersons, or make composite indices based on secondary statistical information collected from a variety of sources. Given the different methodological approaches, as well as the relatively small number of countries covered in the survey, it is only appropriate to draw tentative conclusions from the comparisons. Generally, coefficients of .50 are considered to be moderate, while those closer to .80 are considered strong.[20]

It is also important to emphasize that the different sources of data take different views of governance to the WGS and we would not expect ours to be highly correlated with some measures. Some other data sources only focus on specific aspects of governance and thus comparisons can only be along limited lines. When indices deal with more specific sets of issues than our more general survey of governance, we have tried to disaggregate our governance index and match up relevant items in our index with the specific aspects of governance examined in the other measures. For example, given the nature of the questions, we can only compare a couple of the WGS questions with the political risk indicators generated by ICRG. We have, however, included all the WGS items when we compare to the aggregate governance measures prepared by the World Bank. It is always a challenge to compare different measures based on different definitions and data sources. We have tried to do this in responsible way to demonstrate the performance of the WGS.

The World Bank Measures

A team led by Daniel Kaufmann at the World Bank has constructed an overall governance measure using numerous indicators collected from 14 different sources.[21] The World Bank

---

[20] Hatcher,L and Stephanski, E. A Step by Step Approach to Using the SAS System for Univariate and Multivariate Statistics, SAS Institute, Cry NC. 1994.
[21] Kaufmann, D., Kraay, A., and Zoido-Lobaton, P., 1999a, *Governance Matters*, World Bank Working Paper, Washington D.C.: World Bank. Daniel Kaufmann, A. Kraay, and P. Zoido-Lobaton, 1999b, *Aggregating*

combine these indicators into six different dimensions of governance, namely Voice and Accountability, Political Stability/Lack of Violence, Government Effectiveness, Regulatory Framework, Corruption, and Rule of Law. There is a relatively large overlap with the type of issues the WGS focuses on. When we compare our aggregate WGS governance index with the overall World Bank Indicator, we find a relatively robust .77 correlation. Despite the challenges of comparability, it is certainly reassuring that our overall rating compares relatively well to the only other comprehensive rating of governance.

However, as pointed out, the World Bank's definitions and the indicators included in the six dimensions are different from those used in the WGS. So, to better compare the WGS with the World Bank dimension ratings we looked at the indicators used by the Bank in creating its ratings and created corresponding WGS indicators by including the questions that were the most similar. The construction of these indices is presented below:

- For the Voice and Accountability dimension we selected the following WGS items to include in the WGS Voice and Accountability indicator: 1, 2, 3, 4, 6, 7, 8, and 10.

- In the case of Political Stability/Lack of Violence, we used items 11, 14, and 15.

- For Government Effectiveness we include items 12, 17, 18, 19, and 20 from the WGS.

- In the Regulatory Framework dimension, we had only item 22 that matched up well with the Bank's indicators.

- For Corruption we used items 21 and 23 from the WGS.

- Finally, for the Rule of Law dimension we used WGS items 5, 21, 26, 27, 28, and 30.

The results are presented in Table 8 – and again are generally reassuring. The correlations between the various World Bank dimension and corresponding WGS indices are significant at the .05 level or higher, except for the Regulatory Framework where there was only a single item to match up.

Table 8. World Bank Dimension Scores and Corresponding WGS Measures

Comparing the WGS with World Bank Indicators	
Indicators	**Pearson Correlation Coefficient**
WB and WGS Voice and Accountability Indicators	.77
WB and WGS Political Stability Indicators	.53
WB and WGS Government Effectiveness	.59
WB and WGS Regulatory Framework Indicators	.43*
WB and WGS Rule of Law Indicators	.84
WB and WGS Corruption Indicators	.61
Overall World Bank and WGS Governance Indicators	.77

* Significant only at the .09 level.

Freedom House Index

Each year Freedom House produces a survey of Freedom in the World. This assessment, conducted by outside experts, focuses on political rights and civil liberties and uses a seven-point scale to rank performance. "The survey rates countries and territories based on real world situations caused by state and nongovernmental factors, rather than on governmental intentions or legislation alone. To reach its conclusions, the survey team employs a broad range of sources of information, including foreign and domestic news reports, nongovernmental organization publications, think tank and academic analyses, and individual professional contacts."[22]

The Freedom House Index (FHI) is quite different from our approach. In order to compare our findings with the ratings used in FHI we looked at the checklist Freedom House uses in creating its ratings for Political Rights and Civil Liberties and created corresponding indices by including the WGS questions that were the most similar to the Freedom House checklist.

---

[22] www.freedomhouse.org

- The Freedom House Civil Liberties measure was compared to the summed scores of WGS questions 1, 2, 5, 20, 21, 22, 23, 26 and 27.

- The Freedom House Political Rights indicator was compared to the summed scores of questions 3, 4, 6, 7, 8, 9, 10, 11, 12, 14 and 15.

The results presented in Table 9 suggest that the WGS "civil liberties " items correspond quite well to the Freedom House ratings of Civil Liberties.[23] The findings for Political Rights are not quite as good, but still significant at the .05 level.

Table 9. Freedom House and WGS Comparisons

Comparing the WGS with Freedom House	
Indicators	**Pearson Correlation Coefficient**
Freedom House and WGS Civil Liberties Indicators	.63
Freedom House and WGS Political Right Indicators	.53

International Country Risk Guide

Given that our survey is providing an assessment of the political system as a whole, one could expect that there might be a good fit between our data and some the measures contained in the International Country Risk Guide (ICRG) assessment of the political risks for investors in different countries. While most of the ICRG's focus is very different from that of the WGS, we were able to match some of our questions with two of the ICRG's measures, namely the role of the military and rule of law.

- The ICRG's role of the military indicator was compared with WGS item 14.

- The ICRG's rule of law indicator was compared to the summed scores of items 5, 21, 26, 27, 28, and 30.

---

[23] Data from Freedom House, *Freedom in the World, 2000-2001*. Washington, D.C., Freedom House 2001.

Again, the results presented in Table 10 are impressive, with both correlations quite strong and highly significant at the .01 level.

Table 10. WGS and ICRG Comparisons

WGS with ICRG Indicators	
Indicators	**Pearson Correlation Coefficient**
ICRG and WGS Military Indicators	.70
ICRG and WGS Law Indicators	.65

Polity IV

The Polity IV dataset, housed at the Center for International Development and Conflict Management at the University of Maryland, contains coded annual information on regime and authority characteristics for most independent states in the global system.[24] It provides a measure of the extent of "Autocracy" on a scale of 0–10, autocracy being defined in the extreme as a political system where citizens' participation is sharply restricted, chief executives are selected from within the political elite, and there are few institutional constraints on the exercise of power. It also provides ratings of "Democracy" on a scale of 0–10, with democracy defined as a political system with institutionalized procedures for open and competitive political participation, where chief executives are chosen in competitive elections and where substantial limits are imposed on powers of the chief executive. The combined "Polity" score is calculated by subtracting "Autocracy" from "Democracy" – thus on a score of -10 to +10.

We looked at the checklist used in creating the Polity IV ratings for "Autocracy" from "Democracy" (essentially the same indicators for both) and created corresponding WGS indices by including the WGS questions that were the most similar.

- The Polity IV Democracy indicator was compared to the summed scores of WGS items 1, 2, 3, 4, 6, 7, 8, and 9.

---

[24] See: http://www.cidcm.umd.edu/inscr/polity/index.htm#data

- The Polity IV Autocracy indicator was compared to the summed scores of items 1, 2, 3, 4, 6, 7, 8, and 9.

- For the Polity Comp indicator we used WGS items 1, 2, 3, 4, 6, 7, 8, and 9.

Again, the results, presented in Table 11, are robust. From a detailed comparison it is clear that there is some, but far from extensive, overlap between the WGS and Polity IV. For example, Polity IV is particularly concerned with executive recruitment and restraints. Therefore, these comparisons seem to indicate that Polity IV and WGS are capturing similar issues, but in different ways.

Table 11. WGS and Polity IV Comparisons

Comparing the WGS with Polity IV	
Indicators	**Pearson Correlation Coefficient**
Democracy with WGS Polity	.75
Autocracy with WGS Polity	.68
Polity Comp with WGS Polity	.76

Conclusion

The main conclusion that we want to draw at this point is that our approach to collecting governance data at the national level works.   When comparing the WGS with other with the more established holistic governance indicators developed by the World Bank and Polity IV, the correlation coefficients are quite robust at .77 and .76 respectively. Given the differences in how the concept of governance is conceived, we feel these results are about as high as one can expect. When examining the results of the comparisons between other more area specific governance indicators, we find all but one to be in the moderate to strong range. Across the board, regardless of the definitions used, the WGS governance index and selected items demonstrated high significant correlations with well-known measures of governance and

specific elements of governance. However, unlike the other measures, which operationalize the concept of governance to fit specific programmatic interests, the WGS, with its human rights based approach, offers a broader, more complete view of governance.

The WGS is also novel for comparative work, in that it draws assessments from a cross-section of local experts within each of the survey countries, rather than from outside experts or secondary data. With the help of the WIPs' ratings and qualitative comments, our approach generates both quantitative and very rich qualitative data. It is clear that local WIPs are particularly well suited to evaluate the nature of governance in their counties and assess the significance of changes over time. We believe the experience of the project indicates the ability to generate valid and valuable data that for the first time includes the voices of local WIPs – despite the contested nature of the governance concept and the considerable methodological problems in collecting data on this set of issues.

In sum, we can say that our project was aimed at establishing the viability of a new approach of collecting data on governance issues. Many lessons were learned from the data collection exercise. After a careful scrutiny of the way this survey was conducted and the quality of the data that we obtained, we feel sufficiently confident that we have an interesting data set on governance to share with our readers. In the absence of any governance data based on a cross-section of well-informed persons in developing and transitional societies, we are able to present something that no one else has done so far. It is also important to add here that out of the 929 persons interviewed, only one actually questioned the framework or method that we used. The vast majority of those we surveyed not only supported the idea of making a governance assessment but also commented on the value of its comprehensive nature. The country coordinators also mentioned to us that asking people about governance issues was politically feasible in those countries we had chosen. We realize that there are places where such interviews would be impossible but we are encouraged by the fact that we could conduct this kind of survey in countries such as China, Jordan, and Pakistan.