

III

MEASUREMENT, METHODS, AND MORE

The accurate measurement of the performance of governments depends upon the devising of appropriate indicators capable of capturing from the bottom up what citizens regularly look to their national, provincial, municipal, and village polities to provide. We have rigorously attempted to do such measurement in this Index of African Governance. Our approach builds upon the explanatory papers that were published prior to the first Ibrahim Index of African Governance in 2007.¹ In addition to the five categories, thirteen sub-categories, and fifty-seven sub-sub-categories (or “indicators”) that we are using to measure performance and governance in the 2009 Index, more variables could doubtless provide further calibrations and refinements. But we think that the 2009 Index, and the numbers arrayed in it, present sufficient complexity to capture and display the attainments of and the differences between governments at any level, anywhere (not just in Africa).

Measuring what political or organizational entities do is not new. Benchmarking and preparing report cards on various kinds of performance is well-established. Indeed, in recent years, with regard to national governments, indices and ranking systems have proliferated. There are happiness, global peace, global integrity, economic freedom, competition, corruption, political freedom, and many other index offerings.² There are a variety of national, regional, and international attitudinal surveys, some extremely ambitious. But what makes this 2009 Index of African Governance unique (more so than the 2007 and 2008 Ibrahim Indexes) is its attempt to be comprehensive across a broad range of data for all fifty-three African countries, over multiple years.

Among projects that seek to measure governance comprehensively, only the World Bank’s Worldwide Governance Indicators is as complete in its coverage of countries. Other projects with similarly complete country coverage—such as Transparency International’s Corruption Perceptions Index; the UNDP’s *Human Development Report*; and Freedom House’s Freedom in the World Report—are designed to measure *components* of good governance (here corruption, human development, and respect for political rights and civil liberties, respectively). Among other broader projects on governance in Africa, for instance, the United Nations Economic Commission for Africa’s *African Governance Report* (AGR) covered twenty-seven African countries in its first report in 2005 and thirty-five countries in its second report, published in 2009. The latest round of Afrobarometer surveys, conducted during 2008 and released in May 2009, covered nineteen African countries. Twenty-nine African countries have acceded to the African Peer Review Mechanism (APRM), an important self-monitoring governance endeavor.³ However, country review reports are currently available for only seven countries.⁴

1 In particular, see Robert I. Rotberg, “Improving Governance in the World: Creating a Measuring and Ranking System,” in Rotberg and Deborah West, *The Good Governance Problem: Doing Something About It*, WPF Report 29 (Cambridge, MA, 2004).

2 The UNDP Oslo Governance Centre’s “Governance Assessment Portal” will provide a summary of many initiatives. See www.gaportal.org/.

3 As of June 2009. Cape Verde “...formally accede[d] to the APRM at the 11th Summit as the 29th member of the APRM...” See “Communique Issued at the End of the Eleventh Summit of Heads of State and Government Participating in the African Peer Review Mechanism [APR Forum], 30 June 2009, Sirte, Libya,” available at http://aprm.krazyboyz.co.za/index.php?option=com_aprm_about&Itemid=30&page=aprm-forum&nid=2 (last accessed 4 September 2009).

4 See www.aprm-international.org/.

National Sources

The availability of good data drive each of these efforts, not least our own. In preparing the first Ibrahim Index of African Governance in 2007, we found numbers for many indicators much harder to obtain than we expected. Although some figure for some year for most countries is generally available for most of our indicators, obtaining data for every appropriate year from international sources is especially difficult. In order to supplement internationally available sources (such as the World Bank's World Development Indicators, UNESCO, WHO, and so on) and to determine whether some figures are collected in-country, but just not made available internationally, we have been attempting since 2007 to collect good numbers for approximately twenty indicators from national sources in each of sub-Saharan Africa's forty-eight countries. (We have not had research affiliates in the North African countries because we added assessment of them only in this year. In addition, international sources generally have more complete information for North African countries than for sub-Saharan countries, so the need for in-country affiliates has been less pressing. Similarly, we have not focused on finding a research affiliate in Mauritius, a country for which international statistics are relatively complete, as is the information available through the reports and website of the Central Statistics Office in Port Louis.)

In-country research affiliates, directed by the Index staff at Harvard, have sought to gather information from national statistical offices and from ministries of justice, health, education, and so on. This massive effort has been mostly successful; in this year's Index we provide measures that are fuller and stronger than in 2007 and 2008 because of the continued deployment of locally derived numbers. Nevertheless, missing numbers still remain a major problem for a few key indicators and countries. Each of those gaps is mentioned in the notes to the individual indicators.

With data arriving directly from individual countries, there is the added challenge of assessing data comparability across countries. For this reason, we have not been able to use in this year's Index all of the numbers that our research affiliates ably collected. Maternal mortality figures provide a good example of why. As statistics on maternal mortality published in UNICEF's *The State of the World's Children 2009* report illustrate, there are often major differences between country-reported and "adjusted" (and internationally-comparable) figures on the maternal mortality ratio.⁵ Such differences may be due to the varying methods that countries use to measure maternal mortality, as well as, as UNICEF notes, to "well-documented problems of misclassification and underreporting."⁶ In Guinea-Bissau, for instance, the nationally reported figure is 410 deaths per 100,000 live births and the adjusted figure is 1,100.

The interpretation of official national and international crime statistics poses a particular challenge. Variation in official crime rates may reflect not only variation in the actual number of crimes committed, but also variation in record-keeping and reporting—itself an indicator of public faith or lack of faith in a national criminal justice system. Higher official crime rates might thus reflect both a negative and a positive governance outcome. We do not have solutions to these and similar data problems, but we remain attentive to them as we continue to develop more sophisticated methods of data collection, both internationally and locally, through our research affiliates on the ground in Africa.

The use of local data for particular indicators is discussed in the relevant descriptive notes. These notes detail how and when we decide to use local numbers. In several cases when the local numbers are not used directly in the Index, we summarize the data collected as a point of comparison and so that other researchers may use these figures.

The Latest Numbers

In each edition of the Index, assessment is based on the latest numbers available. Given the type of data that we use (hard numbers, focused on outcomes), the latest numbers available for most indicators for most countries are for two

5 See WHO, UNICEF, UNFPA, and The World Bank (prepared by Lale Say and Mie Inoue of WHO, and Samuel Mills and Emi Suzuki of The World Bank), *Maternal Mortality in 2005: Estimates Developed by WHO, UNICEF, UNFPA, and The World Bank* (Geneva, 2007).

6 UNICEF, *The State of the World's Children 2009: Maternal and Newborn Health* (New York, 2008), 149.

years prior. Thus, the 2007 Ibrahim Index was up-to-date through 2005, the 2008 Ibrahim Index through 2006, and the 2009 Index through 2007. Other projects that rely on similar types of figures show a similar time lag (see, for instance, the well-known Human Development Index).

In accordance with standard research practices, the numbers used in our Index are fully updated in each year—both for the latest year *and* retrospectively for previous years. Thus the 2009 edition of the Index of African Governance reflects the best available numbers for all years, as of approximately June/July 2009.⁷ Although not all indices submit to this rigorous standard, we do so in order to take advantage of all the advances in research made during each year, while producing an Index from which meaningful comparisons can be made over time. Data sources improve continuously due to new information, better estimation techniques, and corrections of simple errors. Standard international sources on even the most well-used indicators, such as GDP and inflation, are regularly revised. In several cases, new sources have been published that measure specific indicators more accurately than previously available sources.

Retrospective revision—compelled by international data reassessments and the importance of incorporating newly available information—means that some numbers used in the 2009 Index of African Governance are different from those used in our 2007 and 2008 Ibrahim Indexes. In other words, a country's rank in the 2008 Ibrahim Index expressed our assessment, based on the best information available at that time, of its governmental performance relative to other countries in that year. If better information became available in 2009, that earlier assessment has been adjusted. Such revisions do imply some changes in previous years' overall scores and rankings. Comparisons year to year should thus be made using only the data published in each year's Index edition, as is standard practice.

Normalizing the Data

This Index avoids being prescriptive in terms of policies, letting the numbers tell the performance story. Those interested may review the raw data for each of the fifty-seven indicators in order to develop a full, nuanced picture of performance in each country. But, the makers of the Index also agreed that a single composite score for each country—and, based on that, a ranking of all countries—was important for broad comparisons.

In calculating this composite score, we have had to make several key decisions. The first was how to normalize the raw data, putting it on a common scale so that the many different measures included in the Index could be compared and combined to calculate a single overall score.

There is no single standard method of calculating an index.⁸ Deciding among methods depends upon various considerations, including the type of comparisons that the analyst seeks to emphasize, the characteristics of the underlying data, and the theoretical value of placing emphasis on “outliers.” It was critical for this Index to emphasize relative simplicity as compared to other methods (i.e., the ease with which the results could be understood by non-statisticians).⁹ The makers of the Index also sought to ensure that the real data themselves—rather than targets or reference points set arbitrarily—governed the scaling of the data.¹⁰ Doing so was considered important for the Index because of the theoretical difficulties and potentially random nature of selecting constant reference points for all fifty-seven indicators. Additionally, we sought a method that would not discard information, when available, about variation

7 Exact dates on which our sources were last accessed are noted in the descriptions for each indicator.

8 For a useful summary of methods, see Michela Nardo, Michaela Saisana, Andrea Saltelli, Stefano Tarantola, Anders Hoffman, and Enrico Giovannini, “Handbook on Constructing Composite Indicators: Methodology and User Guide,” OECD Statistics Working Paper, (Paris, 2005).

9 This was one reason that z-scores were not used.

10 This was one reason that “distance from a reference (or target)” methods were not used.

among countries. For instance, although some of the indicators contain just three possible values (such as “not at all,” “partially,” and “fully”), others have a value from 0 to 100. Because the additional variation when available is useful in distinguishing among countries (even if not available for all indicators), we refrain from normalizing the indicator data by grouping values into three (or five, or ten) categories.¹¹

Prior to the publication of the first Ibrahim Index in 2007, the Index authors examined several methods. For a description of the pros and cons of these methods, see the “Measurement, Methods, and More” essay in our 2008 report. The Index method employed since the 2007 Ibrahim Index is as follows:

For each indicator, the raw data are re-scaled such that the minimum value across all years of the Index (2000, 2002, 2005, 2006, and 2007) receives a score of “0” and the maximum value across all years of the Index, a score of “100.” For each indicator in each country in each year, the score is calculated as follows:

$$SSC_c^t = 100 \times \left(\frac{x_c^t - MIN(X)}{MAX(X) - MIN(X)} \right), \text{ where } x_c^t \text{ is the raw value for that indicator for country } c \text{ in year } t \text{ and } X$$

describes all raw values across all countries for that indicator across all years 2000, 2002, 2005, 2006, and 2007.

(Note that because high values may indicate good performance for some indicators and low values good performance for others, we subtract this sum from 100, as appropriate, so that the best performers always receive the highest values and the worst performers the lowest values. Details about scaling particular indicators can be found in the descriptive notes to each and in the introductory notes to the various categories.)

The key benefit of this method is that scores for each country can be compared over time for 2000, 2002, 2005, 2006, and 2007, as well as relative to other countries within the same year. In addition, the real data for each indicator determine the minimum and maximum values. Doing so is useful for indicators such as “battle-deaths,” where it is difficult to make predictions about the maximum possible values. An argument can also be made about its use for other indicators for which the possible range of values is more predictable, such as the literacy rate. For instance, for the literacy rate indicator, one might set the minimum possible value at 0 percent and the maximum possible value at 100 percent. However, because the real data occupy a smaller range for this indicator (17.1 to 91.8 percent), arbitrarily setting the scale at 0 to 100 percent will mean that country values for this indicator will be lower than for other indicators that do not have this characteristic. That fact might critically affect sub-category, category, and overall scores adversely; therefore, we refrain from scaling in that manner.

The key drawback to this method is that the scores of the Index of African Governance must be adjusted in each year—both for the latest year and previous years—when there are changes in the real minimum and maximum values in various indicators. However, as noted above, the underlying data are also adjusted annually; thus, some such changes are inevitable.

Weights and Weighting

A second key decision in the calculation of the Index involved weighting. Within the five main categories of the Index of African Governance (Safety and Security, Human Development, and so on) we have had to decide how to weight the sub-categories—i.e., whether the sub-categories were to be counted equally or by some other method to arrive at a total score for each country for each category. We have also had to decide the weighting of the indicators within each sub-category. Our basic rule is equal weighting, which reflects the equal importance of each category and sub-category in the Index’s conceptualization of governance: In other words, within each category, the scores of each sub-category are

¹¹ This was one reason that methods employing rankings only or categorical scales were not used.

weighted equally and the sub-category scores are averaged to arrive at the overall score. Within each sub-category, each sub-sub-category (indicator) is weighted equally and the indicator scores are averaged to create the sub-category score. (See “The Meaning of Governance” for summary tables of the Index categories, sub-categories, and indicators.)

There are two exceptions to this rule: First, in the category of Safety and Security, we would ideally have counted the two sub-categories of that category equally, as both National Security and Public Safety are key components of that political good. But, after reflecting at length on the quality of the data and the extreme problem of missing data for crime (explained in the introduction to Safety and Security), we decided that it was more fair and more accurate to weight National Security (where the data were comparatively complete) twice as much as Public Safety, thus weighting the first two-thirds and the second one-third.

Second, in the sub-category of Respect for Civil and Political Rights (in the category of Participation and Human Rights), we include an indicator of women’s rights that is based on the sum of three other indicators (“sub-”indicators) for women’s economic rights, women’s political rights, and women’s social rights (each drawn from the Cingranelli-Richards Human Rights Dataset, as explained in the descriptive note on women’s rights). Given the importance of these three sub-indicators, we list the results for each separately in the Index, just as we do for the other indicators.

In deciding to weight categories equally, we considered several alternatives. One of the makers of the Index long assumed that Safety and Security should be over-weighted in any final result because positive governmental performance is impossible without a large measure of security. The Index’s Executive Council, comprised of distinguished African scholars and practitioners, urged us to weight each category equally on fairness grounds, although one or more of its members favored over-weighting Safety and Security and/or Sustainable Economic Opportunity and Human Development.

In the final analysis, we chose to weight all categories equally—a decision based on our underlying conceptualization of governance. In order to assess the effect of this decision on our results, we also ran the numbers for the 2007 Ibrahim Index in several different ways, over-weighting Safety and Security, Sustainable Economic Opportunity, and Human Development separately. This analysis suggested that our chosen method was generally robust to these different weighting schemes. Although there were some variations among middle-ranking countries, the best and worst performers were similar across these methods. For that reason, and because—theoretically—we are persuaded that weighting each category equally is fair, and not challengeable on theoretical or epistemological grounds, the ultimate ranking scores that we present are not weighted to favor one category over any other.

As our work on the Index has matured since 2007, we have completed additional analyses of the Index data to test the robustness of our results for different conceptualizations of governance and weighting schemes. As part of these analyses, we have explored how countries might rank against each other if governance were measured in a narrower, more traditional way, such as to include only our categories of the “Rule of Law, Transparency, and Corruption” and “Participation and Human Rights.” (In other words, what happens if these two categories are weighted 50 percent each, and the other three categories are given no weight at all?) Rankings based on this approach are given on page 23.

Missing Values

The regular collection of reliable statistics, such as those used in our Index, is expensive and labor-intensive. Not surprisingly, missing data is a major problem for all analysts of African governance. A final, related decision for the Index authors thus concerned how to represent missing values. Our solutions here have again been governed by the project’s commitment to providing ratings that are transparent and accessible to citizens, civil society, and governments, as well as to other scholars—statisticians and non-statisticians alike. Thus, although we are aware of other more statistically complex methods of imputation, we have chosen the simplest methods where possible.

The Index authors have continued to evaluate the effects of these methodological choices on the overall rankings each year. This year, we have had the opportunity to do so in light of the findings of a sensitivity analysis of the 2008 Ibrahim Index published by Saisana, Annoni, and Nardo (2009; see excerpts in this report). The report concludes that the method adopted in our 2008 Ibrahim Index is robust overall. The authors have made one major change to the Index this year as a result of this analysis: we have calculated the Index without one of its original indicators, the Environmental Performance Index. (Alternate results are also presented with this indicator included.) Although this indicator is important, we have chosen to not include it in the Index due to the findings of the sensitivity analysis and our continuing concerns over missing data problems with this indicator. Saisana, Annoni, and Nardo also make several other recommendations for missing data, including the use of the “hot deck” method for imputing missing values. The authors are still analyzing the implications of such changes and thus the findings of this sensitivity analysis will be further incorporated into the next, 2010, Index of African Governance.

There are three types of missing data in the Index. First, for some indicators, such as the homicide rate, values are simply unavailable for almost all countries for all years. In these cases, we have either found “second best” measures for essential indicators, or have not included these indicators in our Index for the time being. In the case of crime, for instance, we include a measure of the “level of crime,” rather than actual crime rates per 100,000 people. On the other hand, we have chosen not to include an indicator on tertiary education rates, in addition to those already included on primary and secondary schooling. Details about specific indicators are discussed in the descriptive notes for the Index categories and indicators.

Second, missing data are a particular problem for some countries. The extreme example in the Index is Somalia, particularly in the categories of Sustainable Economic Opportunity and Human Development. We have chosen to keep Somalia in the Index given the project’s commitment to assessing all African countries. In addition, we find its ranking at the bottom to be accurate. However, we caution readers about drawing conclusions about governance performance in Somalia based on changes in its Index scores, which are based on highly incomplete figures. On the other hand, the problem of missing data was so extreme for two countries, Western Sahara and Somaliland, that we chose not to include them in the 2009 Index. Although both Western Sahara and Somaliland act as states and Western Sahara is a member of the African Union, both lack full international recognition. Thus, many standard international statistics are simply not available, and we are unable to rate them in a way that is comparable to other African countries.

Third, for some indicators, values are available for some or selected years, but not for every year for every country. For instance, it is common for figures on poverty rates to be collected only every ten years. In addition, some projects only began collecting data at some point after 2000. The World Bank’s *Doing Business* surveys, for instance, were first published in 2004. Prior to 2004, there is no source for some of the indicators collected by the *Doing Business* project. *Doing Business* also only began collecting statistics on some countries (such as Liberia) after 2004. For this type of missing data, we have either used the closest year’s figure as an estimate (the year prior, if available), or we have calculated the Index without the figure. For instance, we use an indicator from the *Doing Business* surveys on the number of days necessary to start a business. Data are not available for the Index years 2000 and 2002, and we treat the value corresponding to 2003 as an estimate for these years. Given that we do not want to exclude this indicator for the years 2000 and 2002, that no earlier values are available, and that available values seem relatively stable year to year, we think our approach is reasonable.

In several instances (particularly in the category of Human Development), however, even such rough estimates are unavailable for some indicators for some countries in some years. In these instances, we have calculated the sub-category, category, and overall Index scores without these missing data points, thus calculating the relevant sub-category scores by equally weighting the scores from the available indicators (i.e. mean substitution). For instance, the poverty sub-category is calculated from three indicators: the rate of poverty using the national poverty line, the rate of poverty using the international (\$1.25 per person per day, PPP) line, and inequality measured in terms of the Gini index. Data are available for São Tomé and Príncipe on poverty at the national line and on inequality, but not on poverty at the international line. São Tomé and Príncipe’s score is thus calculated based on an average of the scores for the two available

indicators. Effectively, this approach is the same as estimating missing values to be the average of available indicator scores in a respective sub-category. The resulting sub-category score would be the same if we were to estimate São Tomé and Príncipe's score on poverty at the international line as equal to the average of its scores on other indicators in this sub-category. We have carefully noted throughout the Index each instance of missing data and any estimates used.

This method should yield reliable results if the real values for these missing data points are similar to those for other indicators in the same sub-category relative to other countries. However, if there is significant variation in a country's performance across indicators in some sub-categories, this method will produce scores that are either lower or higher than reality, depending on the case. In the absence of better information about all of the instances of missing data, however, the method employed reflects our best overall assessment at this time.

Statistical Quality

A final challenge related to missing data is statistical quality. Often, the quality of statistics available for countries and the countries' quality of governance as ranked in the Index go together. Thus, the exact country scores at the low end of the Index should be regarded with caution, although relative positions are informative. The fact that so little information is available about some countries is perhaps not surprising: Yet it is unclear how a government can govern effectively if it is unable or unwilling to collect and make public basic information about the well-being of its citizens. If it does not even collect the sort of information used in this Index, on what basis does it monitor the effects of its policies or draw up strategic plans?

In addition, country experts will surely have questions about many of the specific numbers used in this Index. Are literacy rates in Zimbabwe really that high? Do so many Malawians really have access to potable water? Is the ratio of students to teachers in the Democratic Republic of the Congo really so comparatively low? The authors of the Index continue to wrestle with such challenges, inviting comments from our readers and searching for new and better data sources. However, in large part, the quality of available statistics is a problem far beyond the capacity of the Index project: we cannot, and should not, recreate the work of national statistical agencies and international organizations charged with the collection and dissemination of statistics on population demographics, health, education, infrastructure, and so on.

Sub-regional Performance

A methodological challenge that we will seek to address more fully in future years has to do with variation in governmental performance within countries. In some cases, such variation can be extreme. For example, the difference in governance and performance between Somalia (a collapsed state with older, if out-of-date, borders) and Somaliland (a largely unrecognized but functioning state confined to the territory of pre-1960 British Somaliland) cannot be parsed fully in this Index. Few international sources present data that recognize this meaningful distinction, and hence, this Index cannot evaluate Somaliland in a manner analogous to other, internationally recognized countries. This year, however, the Index team has collected significant information on Somaliland through research in Hargeisa and is working to provide an estimate of Somaliland's performance, which will be published on our website.

More generally, in regionally or ethnically divided countries (for instance, the Sudan) the quality of governance has clearly differed markedly across (ethno-) regions, and recent national-level statistics may be based on censuses or surveys conducted in only one region. Such variation is similarly masked in most projects that rely on national-level data. The Index team thus encourages sub-national studies employing the Index method. While the Index of African Governance does not currently include sub-national evaluations, it continues to work with local scholars and organizations to develop such studies, country by country.