SGI Sustainable Governance Indicators 2011

Designing Sustainable Governance Indicators 2011: Criteria and Methodology

Martin Brusis, Jörg Siegmund

BertelsmannStiftung

Designing Sustainable Governance Indicators 2011: Criteria and Methodology

Martin Brusis, Jörg Siegmund

Martin Brusis, University of Munich

Jörg Siegmund, University of Munich

This chapter explains the concepts, assessment criteria, and methodology underlying the Bertelsmann Stiftung's Sustainable Governance Indicators (SGI) 2011, which evaluate and rank sustainable governance in 31 member states of the OECD. Although the conceptual architecture of the Sustainable Governance Indicators remains fundamentally unchanged, a few methodological and operational changes have been introduced for this edition. These changes are based on the results of a thorough internal and external evaluation of the SGI project that took place following the publication of the 2009 edition. The key changes include:

- Changes to the Status Index: The criterion assessing the integration of migrants (S14) is now subsumed under the category of "Social Affairs" instead of "Security." In addition, basic socioeconomic parameters have been integrated primarily into the category of "Economy and Employment," and are no longer accorded their own separate criterion.
- Changes to the Management Index: For the 2011edition, we have streamlined the categories in the Management Index that assess the strategic capacity of executives. The "Institutional Learning" category has been broadened to include the adaptability of governments to international developments, which was previously measured as a separate category. Whether and how governments anticipate veto players in the legislative process is now assessed through criteria assessing effective implementation (M6), rather than attempting to ascertain this through a numerical evaluation of lawmaking.
- New procedure in creating country reports: Two instead of three experts now analyze and assess the situation in the respective OECD member states under review. By beginning with a single country report written by one expert and subsequently revised by a second expert, we take into account that most expert assessments provided for the SGI are not diametrically opposed, but complement and contextualize each other. The numerical scores provided by both are done so independently. Since the two country experts communicate in producing the report, reports and ratings tend to become more reliable.
- Modified aggregation: The disaggregate ratings of experts are no longer standardized before being aggregated. This modification, by emphasizing the position of a country relative to the response options in the questionnaire, links numerical to verbal assessments and improves transparency in the aggregation process.

Our starting-point in developing the SGI has been the acknowledgment that all developed industrial countries face several major reform challenges at the outset of the twenty-first century. These challenges include economic globalization, climate change, resource depletion, aging societies, immigration, and new security threats. In order to address these challenges, advanced industrial countries must develop models of what we refer to as "sustainable governance": the political management of public affairs that adopts a long-term view of societal development, takes into account the interests of future generations, and retains capacities for social change.

Policymakers and scholars will naturally have different opinions about how to manage public affairs and provide public goods in a sustainable way. As long as there is no consensus about the best possible set of policies, approaches that measure sustainable governance in terms of the degree to which a given catalogue of reforms has been implemented will be criticized as oversimplified and mechanistic, as well as biased in applying one standard for all. Moreover, executing checklists of externally devised reforms may not necessarily entail nor result in sustainability.

Though less readily tangible, dynamic adaptation, institutional learning, and innovation are more telling indicators of sustainable governance in a given state. For these reasons, we rely on a mechanism of indirect measurement that compares both the policy performance and executive governance of various states. A cross-national comparison of policy outcomes allows the impact of governance models to be assessed *ex post*. Evaluating policy outcomes has become increasingly feasible as the international consensus on "good" policy outcomes, such as high employment rates or low carbon dioxide emissions, has grown. Better policy outcomes or greater improvements—across borders or over time—can be read as indications of a greater capacity for governance in a given country.

At the same time, however, it must be said that an *ex-post* evaluation of policy outcomes sheds little light on the so-called secrets of success or on whether outcomes can be attributed to deliberate strategy, charismatic leadership, favorable circumstances, pure chance or the organizational makeup and acumen of executive actors. The SGI focus on the organization and role of executives because they can be influenced by resolute, targeted policies. Despite widespread acknowledgment that the ways in which executive actors govern are important (Andeweg 2003), the governance of executives has yet to receive sufficient scholarly and public attention. Executive actors have command of significant resources and the power to translate popular preferences into policies. How these resources and power are managed may not guarantee the success of governance, but it surely affects the chances for policies to succeed. Thus, focusing on executives can add an important *ex-ante* perspective on sustainable governance.

The concept of "executive governance" refers to more than the institutional arrangements of governing. The concept also comprises the mechanisms for and patterns of interaction between the core executive and its organizational environment (i.e., within the executive itself and in the wider political system). In short, it is the *institutional capacity* for governance that is compared rather than the individual decisions of prime ministers or the leadership attributes of charismatic reformers.

This kind of institutional analysis raises the question of how we can know whether a country's established model of executive governance is more or less sustainable. Although scholars broadly agree on the merits of democracy, they continue to argue over which institutional arrangements are superior: presidential or parliamentary systems of government, majoritarian or consensual democracies, and small or large public sectors. These macro-level categories of analysis refer to institutional features deeply embedded within a given country's institutional culture and tradition. As such, they constitute the given framework or environment in which capacity-building reformers must act, but which often lie beyond the scope of their discretion.

In contrast, the micro-level functions and processes required to run a government have increasingly been subjected to cross-national evaluation, transfer, and learning. Driven by a growing interest in "good governance" and "performance management," political practitioners and scholars alike have been developing a body of best practices in managing political processes. International organizations and agencies—such as the Organization for Economic Cooperation and Development (OECD), the World Bank and the European Commission—have used these best practices as reference points for processes involving benchmarking and peer review (Ben-Gera 2004; Nunberg 2000; OECD 2005). This evolving know-how regarding international public management covers areas such as strategic planning, interministerial coordination, the drafting of

legal acts, monitoring, budgeting, auditing, task delegation, institutional learning devices, and public communication and consultation policies.

Drawing on this shared knowledge, we assume that the extent to which a government has established best practices in performing its functions can be taken as an indication of normatively "superior" governance that is likely to prove more sustainable. It should be noted that an executive with greater institutional capacity does not necessarily generate improved socioeconomic performance or better performance in terms of policy outcomes or the quality of democracy, all of which can be influenced by the factors previously mentioned (e.g., favorable circumstances, charismatic leadership and so forth). Nevertheless, greater institutional capacity does improve the likelihood that political leaders will make decisions that can fully harness a country's potential and maximize its performance. Furthermore, a micro-level evaluation of executive-level governance pinpoints observable deficiencies and may allow efforts toward improving the quality of governance to be guided in a more targeted way.

Concepts, questions, and indicators

For the reasons explained above, the SGI incorporate a two-tiered system of measuring reform capacity that assesses both governance outcomes and the institutional potential for sustainable governance. These two distinct aspects are represented in two composite indicators: a Status Index and a Management Index. These two indices consist of 147 individual items—100 and 47, respectively. Eighty-two quantitative indicators are derived from information collected from public data sources, whereas experts for each country have provided 65 qualitative assessments. (For a detailed list of indicators and questions, please see the appendix at the end of this volume as well as the SGI 2011 Website www.sgi-network.org.)

Table 1: Composition of the Status and Management Indices

	Status Index	**Management Index**
Dimensions	2	2
Categories	4	6
Criteria	19	13
Indicators/Items	100	47
***Of which*:** **Expert assessments**	29	36
Quantitative indicators	71	11

The Status Index

The Status Index reflects the growing political and scholarly consensus on what good policy outcomes entail as well as the importance of a high-quality democracy as a framework for policy performance. The SGI's concept of democracy—which is manifest in the first dimension of the Status Index—includes not only the rights of political participation and electoral competition, but also the rule of law (Merkel 2004). Since all OECD member countries are democracies, the SGI's questions in this category focus on the quality rather than the presence of democracy. There are a series of questions designed to address whether citizens face discrimination in the electoral process, how citizens can access public information, the degree to which the media are independent and diversified, how well states protect civil rights, and whether the government and administration act predictably and in accordance with the law (criteria S1–S4).

In order to assess policy performance—which is manifest in the Status Index's second dimension—four broad policy areas are examined. These areas constitute political priorities for governments in responding to the key challenges that most advanced industrial states face. These challenges are: the global integration of markets and its effects for national economies and competitiveness; aging societies and their effects for the sustainability of social security and pension systems; new security risks, such as those arising from terrorism and transnational crime; and the depletion of natural resources resulting from global climate change as well as the scarcity of human capital.

In order to address these challenges appropriately, governments must devise and coordinate various inter-related policies that can be grouped into four broad policy areas: economy and employment; social affairs; security; and resources. In the Status Index, each policy area constitutes a category that is comprised of several criteria, as follows:

(1) Economy and Employment: Economy, Labor Market, Enterprises, Taxes, Budgets (criteria S5–S9);

(2) Social Affairs: Health Care, Social Inclusion, Families, Pensions, Integration (criteria S10–S14);

(3) Security: External Security, Internal Security (criteria S15–S16);

(4) Resources: Environment, Research and Innovation, Education (criteria S17–S19).

Each policy area is evaluated by the experts for each country, who draw upon quantitative indicators from public data sources. The experts were asked to evaluate the extent to which a particular policy realizes specified objectives, such as the goal of fiscal sustainability in the case of budgetary policy (criterion S9). These objectives have been carefully selected and defined in order to preclude ideological bias, and to make sure that they would be broadly accepted and supported by citizens, policymakers, and scholars across both political and value-based divisions.

For example, the objective of family policy (criterion S12) is stipulated as enabling women to combine parenting with participation in the labor market. The question is formulated so as not to suggest a preference for either the traditional single-wage-earner family model or the so-called working mother model. Instead, the question presupposes that an optimal system of family support enables women to decide individually whether and when they would prefer to remain full-time mothers or to work full- or part-time.

The question regarding health care policy offers another example (criterion S10). This question avoids any biased leanings toward either a predominantly private or public health care system.

Instead, the objective of public health care policies is defined as providing high-quality health care for the largest possible proportion of the population and at the lowest possible costs. In an explanation appended to the question, experts are instructed to apportion less weight to the criterion of cost efficiency if the criteria of quality and inclusiveness can be considered fulfilled.

In selecting the performance indicators from public data, we have been careful to choose those indicators that are clear in meaning, do not invite ambiguous interpretations and are available for all OECD countries. We have also sought to avoid including model-specific indicators that might be seen as being biased in favor of particular types of economies. For example, the public sector's proportional contribution to GDP is not used as a performance indicator, because doing so would entail using a disputed interpretation as well as implying a bias in favor of either liberal, Anglo-American market economies with small public sectors or of social-democratic Scandinavian market economies with large public sectors. Moreover, the scholarly debate about the varieties of capitalism shows that there is no consensus about which institutional model is the most sustainable and that distinct models still persist in spite of all tendencies toward convergence (Hall and Soskice 2001; Howell 2003). Adopting such a distinction between two or three institutional models would not provide an appropriate basis for an evaluation because doing so would conceptually "freeze" certain features and ignore the dynamics of change, especially when it comes to continental European models.

As a whole, the SGI aim to provide a composite image of a state's performance across various policy areas. As a result, some indicators refer to policy outputs rather than the impact of public policies on society (or "outcomes"). An example of an output indicator would be "Pre-primary education" (indicator S19.6), which addresses public spending on pre-primary education as a percentage of GDP. Together with other outcome indicators, such as "Upper secondary attainment" (indicator S19.2), which identifies the share of the population with at least upper secondary attainment, this output indicator is considered sufficiently unambiguous as a descriptive measure of the success of education policy. Merging output and outcome indicators does not deny the existence of causal links between indicators, such as those between spending on pre-primary education and PISA results (indicator S19.5). At the same time, however, from a conceptual standpoint, the SGI are less concerned with such links than with the relationship between policy performance and executive governance.

The Status Index also includes a few indicators that describe changes over time rather than levels, such as the inflation rate (indicator S5.3) and the ratio of increase in private versus public spending on health care (indicator S10.2). These and other indicators of change over time have been selected because they describe important policy effects that complement the performance profiles of states. That these two types of indicators (change and level) are combined should be viewed against the background of the SGI's generic design of assessing both aggregate policy performance and executive governance.

In summary, the Status Index combines democratic and policy performance items because it conceives of high democratic standards as necessary scope conditions for policy performance. Moreover, high-quality performance in both areas can be understood as the result of strategic governance by executive actors who remain accountable for their actions.

The Management Index

The Management Index reflects the consensus practitioners and scholars have developed on what good practices in governance entail. The index first examines the extent to which core executives act strategically and can rely on institutional capacities for strategic policy-making. This dimension, labeled “Executive Capacity,” is based on a commonly accepted notion of governing that identifies the government or core executive as the key actor in governance (criteria M1–M8) (Knack, Kugler and Manning 2003; Weaver and Rockman 1993). The Management Index then analyzes the role of actors outside the executive and the extent to which these actors hold governments accountable, enhance the knowledge base of decisions, and deliberate the normative appropriateness of policy decisions. This dimension of “Executive Accountability” reflects the degree of importance attained in governance by actors outside the executive (criteria M9–M13) (Pierre and Peters 2005).

In their theoretical account of governing, Pierre and Peters describe the state’s dependence on these actors as follows: “states must be open to a wide range of information, including much that is uncomfortable and dissonant, if it is to be successful in governing. In other words, states must be in close contact with the society and utilize social information openly and accurately when governing. This further implies that the state is likely to be in close communication with societal actors who possess much of the information that would be required for effective governing and also generally that the state must be willing to engage in a formal or informal exchange of power over decisions for that information” (Pierre and Peters 2005, 46).

Both dimensions of Executive Capacity and Executive Accountability are further structured into categories and criteria. Three separate components of Executive Capacity are distinguished: Policy Preparation (steering capability), Policy Implementation, and Institutional Learning. These components, in turn, refer to stages in the cycle of policy formation as well as to ideas from the literature on Europeanization, globalization, and policy learning (Common 2004; Dolowitz and Marsh 2000; Radaelli 2003; Wiesenthal 1995). The components address the following factors in particular:

- Steering Capability: Strategic Capacity (government planning and expert advice), Inter-ministerial Coordination, Evidence-based Instruments, Societal Consultation and Policy Communication (criteria M1–M5);
- Policy Implementation: Effective Implementation (management of task delegation to ministers, agencies, subnational governments and private actors) (criterion M6);
- Institutional Learning: Adaptability (governmental capacity to adapt to globalization, Europeanization or transnationalization as well as to internationally coordinate joint reform initiatives), Organizational Reform Capacity (governmental capacity to reform institutional arrangements and improve strategic orientation) (criteria M7–M8).

Executive accountability is subdivided into three separate categories corresponding to actors or groups of actors that are considered to be key accountability providers in theories of democracy and governance (Bovens 2007; Pierre and Peters 2005, 46; Schedler 1999, 17; Schmitter 2004). The questions posed here are, To what extent are citizens informed of government policy-making? (criterion M9); Is the legislature capable of evaluating and monitoring the executive? (criteria M10–M11); and Do intermediary organizations (i.e., the media, political parties, interest associations) demonstrate substantive know-how/provide substantive information? (criteria M12–M13).

As was the case with the Status Index, country experts provide evaluations in response to the individual questions of the Management Index. In addition, quantitative data, such as the number of parliamentary committees, are centrally collected by the SGI team from public sources.

In sum, the Management Index assumes that more accountability—in the form of public scrutiny, information channels, and normative deliberation—improves a country's executive governance. This assumption might be challenged as counterintuitive from a conventional approach of governing that tends to view accountability and participation mechanisms as constraints on executive authority.

Governing may be much easier for some governments which benefit from much more conducive structural conditions and which can rely on enabling actor constellations. In order to avoid equating constellations of circumstances that are categorically different, a comparative assessment of executives must take into consideration the constraints posed by veto players, economic crises, hostile popular attitudes, and historical path dependencies. The remainder of this section discusses how the SGI deal with these issues.

Veto players

Veto player theory differentiates between institutional veto players (e.g., a bicameral parliament, constitutional court, etc.) as defined in the constitution and political veto players, which are primarily the different parties that make up a governing coalition but also organized interest groups (Tsebelis 2002). While it is true that a greater number of veto players increases the stability of an adopted policy, stability as such is not necessarily a negative or positive property of governance. Rather, sustainable governance seems to represent a difficult balancing act between decisiveness and resoluteness or the ability to take (and enforce) decisions and to stay committed to them (Cox and McCubbins 2001).

For this reason, the Management Index does not attribute a veto function *a priori* to certain structures of a given political system, such as a large number of (governing) parties, a strong constitutional court or an opposition-controlled second chamber of parliament. Instead, the Management Index treats the influence of veto players as an empirical question to be assessed by country experts and quantitative indicators in various parts of the survey (questions M4.1, M6.1, criteria M10-M13).

In addition, we would argue that veto players do not necessarily block improvements to the status quo. In fact, many authors have noted that veto players might even improve the quality of governance by helping governments to better assess the potential impacts of a given policy. Doing so induces reformers to broaden their bases of support by accommodating veto-player critiques and thereby renders reforms irreversible (cf., e.g., Gehlbach and Malesky 2010; Benz 2003, 230).

We have therefore decided to treat states equally regardless of the number of veto points their governments must deal with or anticipate. One could object to this methodology by arguing that it is easier for a British government to attain a better assessment than a German government, because it effectively has to anticipate fewer veto points. However, governments in systems with more constitutional veto points than the United Kingdom are aware of the additional veto points and therefore can—and, indeed, must—prepare for them accordingly.

Furthermore, a system with few veto points does not automatically receive a good rating for executive governance. Instead, a country's aggregate assessment also takes into account the government's consultations with business and social actors and its communication with the public (questions M4.1, M5.1). For example, a British government that, thanks to its clear legislative majority, is able to pass many laws while ignoring societal interests would receive a poor rating for the respective questions. In comparison, systems with many veto points usually have various social interests represented among their veto players. As a result, governments in these systems that successfully anticipate veto points can generally expect to receive positive ratings for the questions related to public communication and consultation. In other words, these questions function as correctives to questions on effective policy-making that could favor executives in systems with few veto points.

Economic and social distress

The global financial and economic crisis has severely constrained executive capacity in countries under risk of a sovereign default. One could argue that governments in such countries are in a situation that is not comparable with governments in countries with sound public finances. However, there are also scholars who claim that economic and social crises increase executive capacity insofar as they endow newly elected governments with popular mandates for change that aid them in overcoming vested interests (Alesina et al. 2006; Williamson and Haggard 1994). The painful reforms some crisis-ridden governments have been able to implement and the weak correlation between fiscal adjustment and subsequent electoral lossess (Alesina et al. 2010) seem to support this view. Given this ambivalence between possible constraining and enabling effects of crises, an automatic bonus for countries in crisis does not appear to be justified. Rather, we tend to see national governments as the principle agents responsible for—and capable of—coping with the policy challenges during socioeconomic crises and the rigidities of routine politics.

Attitudes among the population

A higher popular willingness to accept costly reforms is likely to facilitate sustainable governance. The assessment of executive accountability in the Management Index takes this contextual condition into account by asking what citizens in a given country know about the motives, objectives, effects, and implication of governmental policies. It is assumed that a deeper understanding of policies among a state's citizens will increase the public acceptance of reforms.

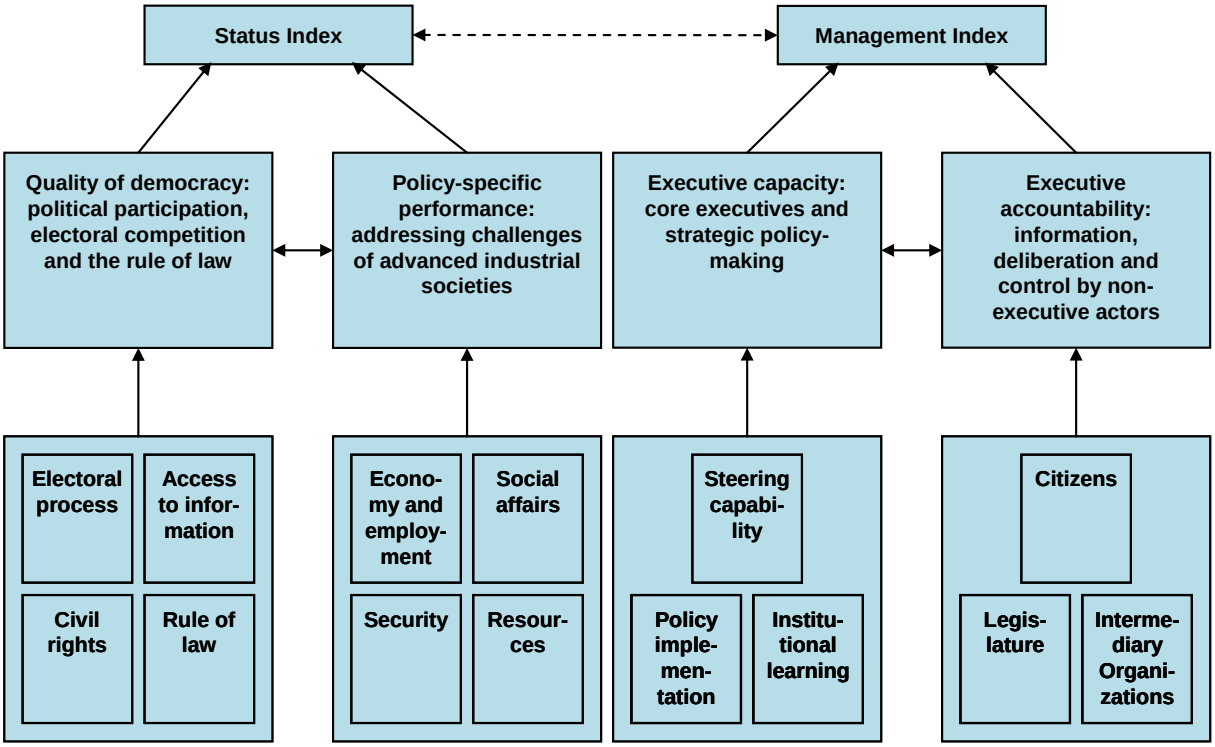
Path dependencies

A country's particular path of historical and institutional development is likely to determine the scope and choice of available policy options (North 1990; Thelen and Steinmo 1992). The design of the SGI encompasses different conceptual and methodological strategies to reflect the formative role of such paths. One strategy, of course, is our approach of indirect measurement. While the Status Index focuses on policy outcomes and avoids evaluating policies on the basis of chosen instruments alone, the Management Index focuses on functions and processes without judging executive governance on the basis of institutional settings that are considered to be superior. Since the conditions under which path dependence constrains or enhances sustainable governance are

not well-known, the approach of indirect measurement applied in the SGI seeks to avoid implying such conditions in both its concepts and questions.

The Management Index also uses another strategy that implicitly addresses the extent to which countries are locked into extant and constraining paths of institutional development. Thus, there are questions under the category of "Institutional Learning" that aim to evaluate whether government actors monitor their own institutional arrangements of governing and improve strategic capacity by means of targeted institutional reforms.

Figure 1: Conceptual tree—Components of the Status and Management Indices



Measurement

To operationalize and measure the concepts used in constructing the SGI, we decided to rely on a combination of statistical data drawn from official sources as well as the qualitative assessments of country experts. Statistical data are generally more reliable than expert opinions, particularly when they are collected by official institutions and by using methods that conform to cross-national standards. At the same time, however, such data often do not adequately cover the full meaning of a concept. We therefore believe that complex concepts can be measured best through the use of expert assessments that take the country-specific context into account and provide "thick" descriptions capturing the nuances of phenomena. Nevertheless, one must always remember that the responses of experts are prone to bias as a result of subjective perceptions and thereby pose problems of intercoder reliability (Munck and Verkuilen 2002).

The SGI's expert survey questionnaire is designed to improve the validity and reliability of expert assessments through the use of six tools and procedural steps. First, many assessment questions are formulated so as to elicit detailed factual evidence rather than broad—and, consequently, more subjective—assessments. In fact, many questions ask for responses that may be cross-checked with responses to other questions, statistical data or data from opinion surveys.

Second, the questionnaire provides detailed explanations of and four tailored response options for each question. This information is intended to illustrate the purpose of a question, to structure the way the expert words his or her assessment, and to provide a standardized framework for the production of the country reports. The experts are instructed to adapt the standardized response options to the individual context of the particular country they are evaluating and to substantiate their ratings (numerical assessment) with evidence in their country report (in the following: "expert report"). The rating scale for each question ranges from one to 10, with one being the worst and 10 being the best. The scale is differentiated by four response options provided for each question. Although the written assessments do not allow for a direct reconstruction of the numerical ratings, they do provide an explanatory background for them.

The questionnaire for SGI 2011 reflects revisions undertaken that draw upon the evaluations of the flagship 2009 survey. These evaluations included reviews by Edeltraud Roller and Steffen Ganghof as well as an intensive debate among all scholars and practitioners involved in the survey. Apart from the above-mentioned changes to the conceptual design, this evaluation process also led us to carefully revise all questions, explanations and response options. The questionnaire is also realized as an online database that contains all information and additional data on the questions. This online tool aids country experts in formulating their assessments while allowing all contributors to keep track of the survey process.

Third, each OECD member state surveyed is examined by two leading scholars with established expertise in the respective countries. To identify subjective bias and reduce any distortion it might cause, the experts were selected so as to represent both domestic and external views as well as the viewpoints of political scientists and economists. One expert writes a draft country report, assessing all questionnaire items. The other expert reviews this report, making comments and providing alternative or complementary content. Both experts are instructed to assess the situation in their countries as of April 2010 and to take into account the period between May 2008 and April 2010 when explaining their evaluation. Although many experts know each other personally, we ensure that their expertise is given independently.

In completing the questionnaire, each expert provides numerical ratings for 65 questions, which means that the evaluations for all 31 countries entail a total of 2,015 ratings (or scores). Whereas the reviewer has access to the written assessments of the first country expert, he or she cannot see the first expert's numerical ratings, which ensures that scores are given independently. For 87 percent of the ratings, the expert scores deviated by two levels or fewer, and for almost a third of them the two country experts chose identical scores. The degree of congruence between authors and reviewers is also reflected in the median standard deviations of their ratings per question, which are zero for four SGI questions, 0.5 for 59 questions and one for only two questions. Since the formulation of a question, explanation, and response options led experts with very different backgrounds to provide very similar ratings, low standard deviations could be interpreted as a measure of precision or reliability.

At the same time, caution should be exercised here with this interpretation, as the motivation behind selecting two experts per country was, indeed, to benefit from the input of a variety of political orientations as well as professional experience. In other words, a high standard deviation is a side effect of the survey's design – not necessarily an indication of a reliability problem. Moreover, the review process eliminated all measurement errors resulting from cases in which obvious misunderstandings produced high standard deviations. Organized as a discursive process between and among the experts and the regional coordinators (see below), the review process allowed the participants to clarify concepts, define the exact meaning of questions and agree on conventions of interpretation that would ensure reliable evaluations.

Fourth, the countries examined by the SGI are subdivided among seven "regional coordinators." These regional coordinators, who are political scientists with both comparativist and area expertise, are each responsible for four or five of the 31 OECD countries, grouped according to their geographical proximity. The regional coordinators monitor the development of the written assessments according to criteria of validity and objectivity, ensuring a fair and balanced country report. In addition, the regional coordinators give numerical ratings based on those provided by the county experts.

Fifth, the regional coordinators review their ratings collectively so as to make it possible to draw comparisons across the entire OECD world. As part of the discussions forming the review process, each regional coordinator is required to explain, defend, and if necessary, recalibrate his ratings and assessments. To make any changes agreed to during the review process more transparent, the coordinators also agree to keep the score within the range of the two country expert ratings. During the review process, six percent of these scores exceeded the range defined by the expert ratings, and each of these deviations was justified in the body of the country reports.

Sixth, as part of a second round of reviews, an advisory body composed of renowned scholars and practitioners that are tasked with making strategic decisions discusses and approves the ratings. This second review resulted in changes to 0.4 percent of the total scores; these moderate adjustments did not alter the proportion of scores exceeding the range of the ratings in the experts' reports.

The other main sources of data for the SGI are quantitative indicators collected from publicly available statistics. Giving country experts and coordinators access to these indicators through the online database allowed them to rely on an equal basis of standardized information. Values missing from public statistics were supplemented with those from previous years or other sources. If the latter was not possible, then the missing value was imputed using the median of the available values. These imputation techniques were applied to ensure the comparability of data drawn from different sources.

Weighting and aggregation

While the expert ratings are based on a unified scale ranging from one to 10, the quantitative indicators are provided using different scales and units of measurement. In order to aggregate the latter into composite indices, the indicators first had to be standardized. This was accomplished by calculating the relative distance from the best performing state and assigning a value to this distance using a scale ranging from one to 10. In cases where lower values of indicators denoted better performance, the scores were inverted so as to guarantee that higher scores always represented better performance. This technique of standardization through a linear transformation was chosen for the SGI because it is both intuitively plausible and easier to understand than, for example, a z-transformation or a transformation based on a logistic function (Matthes and Schröder 2004).

In addition, the chosen method of standardization has desirable effects insofar as it generates scales with identical ranges and fixed end points, limits the influence of outlier values and increases the distance between values lying within a narrow interval so as to emphasize the relative position of states vis-à-vis other states (Giovannini et al. 2005). To check the robustness of our standardization approach, we also calculated the Status Index and Management Index using z-transformed and logistic-function-transformed values. These standardization methods produced very similar rankings, with the scope of rank shifts being limited to a maximum of three ranks with regard to the Status Index. In this case, the large majority of countries ranked the same regardless of the standardization procedure applied. As far as the Management Index is concerned, the scope of rank shifts does not exceed five ranks, and again, more than a third of the countries retained their position no matter which standardization method was used (see appendix to this chapter).

This standardization procedure was modified for the following two quantitative indicators: number of parliamentary committees and average committee size (indicators M10.1–M10.2). As there is no strictly monotonous, linear relationship between the number and size of committees and their ability to monitor the executive, value ranges were determined according to accepted best practices in committee organization (Schnapp and Harfst 2005). For example, countries with 12 to 18 committees and an average number of 13 to 25 deputies per committee were given a score of 10. Scores for countries with more or fewer and larger or smaller committees were depreciated as a result of applying the above-mentioned linear transformation.

For most indicators, there are no broadly agreed-upon, absolute benchmarks that denote top- or bottom-level performance. This is the case either because performance is assumed to increase or decrease continually or because established benchmarks (e.g., the threshold of a general government deficit of three percent of GDP, which the European Union uses as an eligibility criterion for membership in the Economic and Monetary Union) remain contested among scholars and policymakers. For this reason, we decided to define empirical, relative benchmarks by assigning scores of one and 10 to the worst and best performing state, respectively, within the given set of countries.

In contrast, the expert ratings refer to the theoretical benchmarks defined in the questionnaire, and their empirical range was not standardized prior to aggregation. As a consequence, the relative position of a country vis-à-vis other countries matters less for the score than its absolute position on the given scale. This means that top- (or bottom-)performing countries suffer (or benefit) on expert survey items where the empirical distribution does not cover the full scale and clusters close

to the bottom (or top), whereas top or bottom performers on quantitative indicators always score 10 or 1, irrespective of their proximity to other countries. We have chosen to accept the resulting differences in the weights of countries' ranks for individual items because we prefer the ratings to closely reflect the verbal assessments in the reports.

In order to integrate individual items into a composite index, weights have to be assigned to all individual items. Our method of weighting these items has been guided by three considerations: In the first place, we decided that weights should reflect the conceptual status of items, criteria, categories and dimensions that are components of the key SGI concepts of democracy, policy performance and executive governance. Once these concepts were disaggregated into their components, theoretical reasoning was used to identify, define and juxtapose these components. For example, the idea of distinct stages in the policy cycle inspired the disaggregation of the executive capacity dimension into categories, such as policy preparation, implementation, and learning. In contrast, our prior empirical knowledge about, for example, the impact of effective interministerial coordination on the preparation of policies was mainly based on the experiences of practitioners, case-based evidence, intuition, and common sense.

Our knowledge has been particularly limited when it comes to the interaction of individual components with each other. This includes, for example, how interministerial coordination, regulatory impact assessments, and strategic planning jointly affect policy preparation. This uncertainty about effects and interrelations suggests that components might best be considered hypotheses about the presence or fulfillment of a concept (Goertz 2006, 53–58). For example, by defining interministerial coordination as a component of policy preparation, one must assume that effective interministerial coordination improves policy preparation.

On the more aggregate level of SGI categories, it is contended that effective mechanisms of policy preparation in combination with effective implementation and institutional learning increase the strategic capacity of executives. However, we do not know precisely the extent to which individual components contribute to the aggregate concept and whether certain components reinforce or hamper the contributions of other components.

Given these uncertainties, the safest strategy for building indices is to assume, on the one hand, that all components possess equal status as hypotheses about the presence and fulfillment of aggregate concepts and, on the other hand, that each component may partially, but not fully, substitute for the effect of other components. The corollary for the construction of the index at this point is to assign equal weights to all components and choose an additive method of aggregation.

Second, the SGI have been operationalized as a combination of an expert survey and a compilation of so-called hard statistical data. This methodological choice is motivated by taking two facts into consideration: On the one hand, OECD member states are well-charted by numerous datasets, and there are official, cross-national datasets that provide information that is more reliable than subjective assessments of experts. On the other hand, statistical data cover only very specific aspects of more complex realities and ignore a context that can allow for a fuller understanding of an indicator's particular meaning. In order to take this complexity more fully into account, experts were asked to provide contextualized assessments that were then subjected to a review process.

In this way, the combination of expert assessments and statistical indicators assumes that both types of observations have specific strengths and weaknesses, that they cannot fully substitute for

each other, and that neither of them is epistemologically superior to the other (Collier, Brady and Seawright 2004, 252–258). For this reason, we decided to assign equal weight to the expert assessments and the sets of indicators within the policy areas constituting the performance assessment of the Status Index.

Third, the Status Index represents an integrated measurement of the quality of democracy and policy performance in the OECD member states surveyed. Since most OECD member states are stable, functioning democracies, one might be tempted to infer that the assessment of democracy should be given less weight than the performance assessment. However, the results of the SGI expert surveys and other democracy assessments indicate that the quality of democracy varies even between consolidated democracies. Moreover, a high-quality democracy may be viewed as being a necessary framework for strategic policy-making, reform capacity, and meaningful policy performance. For this reason, we have assigned equal weight in the Status Index to the dimensions of Policy-specific Performance and Quality of Democracy.

The Status Index and the Management Index scores are thus derived by calculating the arithmetic means of the scores for their respective two dimensions (i.e., the Status Index's Quality of Democracy and Policy-specific Performance and the Management Index's Executive Capacity and Executive Accountability). The individual dimension scores in the Status Index are derived by calculating the arithmetic means of the criteria scores, which are also derived by calculating the arithmetic means of their respective components. For criteria consisting of expert ratings and quantitative indicators, the two types of items are weighted equally. The categories (i.e., the four broad policy sectors) of the Status Index do not imply a theoretical status within the SGI conceptual framework. They are descriptive and used only to group policy areas. There are therefore no calculations at the category level in the Status Index.

The Management Index's dimension scores likewise represent the arithmetic means of their equally weighted component scores, but the Management Index contains two additional levels of disaggregation so as to reflect the greater diversity of governing practices and mechanisms addressed by the individual questions. The Executive Capacity dimension is disaggregated into stages of the policy process (e.g., preparation, implementation, etc.) that constitute three categories, each of which consists of between one and five criteria that are used to group activities, such as regulatory impact assessment or the organizational reform capacity.

In addition, a distinction is drawn between single items (e.g., with M6.1, which asks about how the government achieves its own policy objectives) and sets of items that are closely related to each other by using lettered annotations (e.g., a, b, c, etc.). For example, intra-executive monitoring mechanisms are viewed as forming such a set consisting of: organizational incentives limiting ministerial self-interest (M6.2a), the monitoring of line ministries (M6.2b) and executive agencies (M6.2c) as well as internal auditing arrangements (M6.2d). These items are weighted equally and, together, are assigned the same weight as M6.1.

The two composite indices—that is, the Status Index and the Management Index— provide scores and ranks for each of the 31 states. The ranking is based on the score that is precise to the second decimal place. If two or more states have the same score at this level of precision, they are ranked equally.

In order to examine the robustness of our chosen approach of weighting, we tested various other weighting models. For the Status Index, for example, we first increased the weight given to the

policy performance dimension. Doing so could theoretically be justified by arguing that most OECD member states are stable democracies where policy delivery matters more than rights and procedures. Second, we explored a model with equally weighted components, thereby weighting the expert ratings according to the total number of items forming a policy area (criterion). Such a model would reflect the assumption that policy areas with more quantitative performance indicators are better charted by hard statistical data and are therefore less dependent on subjective expert opinions. Third (and fourth), we calculated aggregate values based solely upon expert ratings and, conversely, on quantitative indicators. These models are meant to correspond to qualitative or quantitative research designs, respectively.

As Table 2 shows, the aggregate correlations between these models and our default model are fairly high. There are few changes of ranks in models 1 and 2a/2b, which each assign equal weights to all components of a policy area. If the aggregate scores were calculated using expert ratings alone (models 3a/3b), Poland would score much better and Iceland would score worse than in the default model. Ireland and Spain would also fare worse in model 3b, which emphasizes performance over the quality of democracy. More rank changes are generated in the two models 4a/4b, which are based on quantitative indicators exclusively. If the Status Index were composed of “hard” statistical data alone, Germany would suffer the most significant change, falling by six or seven positions. In fact, the Czech Republic would also receive a lower ranking on the “purely quantitative” model and countries such as Austria and Luxembourg would receive a higher ranking.

These results do not fundamentally challenge the ranking of countries in the Status Index, since we have deliberately combined quantitative indicators with expert ratings in order to attenuate biases of both purely quantitative and purely expert-based assessments. However, the comparison with a “purely quantitative” model suggests a cautious interpretation of individual countries’ rank placements in the Status Index is in order.

Table 2: Effects of different weighting models, compared with the chosen model of aggregation

Model	Correlation with default index values	+/–3 ranks	+/–4 ranks	+/–5 ranks	+/–6 ranks	+/–7 ranks
1	.994	Luxembourg/–	–/–	–/–	–/–	–/–
2a	.998	–/–	–/–	–/–	–/–	–/–
2b	.990	–/–	–/Germany	–/–	–/–	–/–
3a	.990	–/Iceland	–/–	–/–	Poland/–	–/–
3b	.979	–/Ireland, Spain	–/Iceland	–/–	Poland/–	–/–
4a	.936	Greece, Japan, New Zealand, South Korea/Belgium, United States	Luxembourg/Poland	Austria/Norway	Iceland/Czech Republic, Germany	–/–
4b	.944	Iceland, Japan, South Korea/ Belgium, Finland, United States	Luxembourg/–	Austria/Poland	–/Czech Republic	–/Germany

Model:

1 = Democracy 30%, performance 70%; expert ratings = indicators

2a = Democracy 50%, performance 50%; components equally weighted

2b = Democracy 30%, performance 70%; components equally weighted

3a = Democracy 50%, performance 50%; only expert ratings weighted

3b = Democracy 30%, performance 70%; only expert ratings weighted

4a = Democracy 50%, performance 50%; only indicators weighted

4b = Democracy 30%, performance 70%; only indicators weighted

In sum, the SGI have been designed to assess reform capacity indirectly by measuring both policy outcomes and executive governance. This indirect approach also entails a degree of methodological, theoretical and political self-restraint in that it assumes that there is no single recipe for reform to be written by social scientists. Moreover, any attempt to communicate the "right" reform will always arouse suspicions of the presence of ideological motives behind the proposed reforms.

Our measurement of executive governance combines state-centric and societal notions of governance. In addition, it focuses on micro-level functions and processes of governance that reflect a growing consensus on best practices beyond traditional macro-categories of political systems. In contradistinction to the existing composite indicators and comparative assessments, the SGI survey explores policy outcomes and governmental practices in greater detail, and it does so for a larger sample of states while using more recent data. Moreover, the SGI indices integrate expert assessments and statistical data in order to combine the advantages of both types of information. As all disaggregate data are published together with the aggregate indices, the data may be customized by recombining them in different ways or with other datasets.

Acknowledgement

The authors are indebted to Margit Kraus (Calculus Consult) for providing important advice and feedback on statistical and technical issues, imputing missing values, exploring different standardization procedures, and constructing Excel sheets for the aggregation of scores.

References

Alesina, Alberto, Dorian Carloni and Giampaolo Lecce. "The Electoral Consequences of Large Fiscal Adjustments, Working Paper, First draft, October 2010, http://www.economics.harvard.edu/faculty/alesina, accessed 22 February 2011.

Alesina, Alberto, Silvia Ardagna and Francesco Trebbi. "Who Adjusts and When? On the Political Economy of Reforms, NBER Working Paper No. 12049, February 2006.

Andeweg, Rudy B. "On Studying Governments". *Governing Europe*, Ed. Jack Hayward and Anand Menon, Oxford 2003. 39–60.

Ben-Gera, Michal. *Co-Ordination at the Centre of Government: The Functions and Organisation of the Government Office*. OECD, SIGMA paper 35, Paris 2004.

Benz, Arthur. "Konstruktive Vetospieler in Mehrebenensystemen". *Die Reformierbarkeit der Demokratie. Innovationen und Blockaden*, Ed. Renate Mayntz and Wolfgang Streeck, Frankfurt/New York 2003. 205–238.

Bovens, Mark. "Analysing and Assessing Accountability: A Conceptual Framework." *European Law Journal*, 13(4) 2007. 447-468.

Collier, David, Henry E. Brady, and Jason Seawright. "Sources of Leverage in Causal Inference: Toward an Alternative View of Methodology". *Rethinking Social Inquiry. Diverse Tools, Shared Standards*, Ed. Henry E. Brady and David Collier, Lanham 2004. 229–265.

Common, Richard. "Organisational Learning in a Political Environment. Improving Policy-Making in UK Government". *Policy Studies* (25) 1 2004. 35–49.

Cox, Gary W., and Matthew McCubbins. "The Institutional Determinants of Economic Policy Outcomes, *Presidents, Parliaments and Policy*, Ed. Stephen Haggard and Matthew McCubbins, Cambridge 2001, 21-63.

Dolowitz, David, and David Marsh. "Learning from Abroad: The Role of Policy Transfer in Contemporary Policy-Making". *Governance* (13) 2000. 5–24.

Gehlbach, Scott and Edmund J. Malesky. "The Contribution of Veto Players to Economic Reform", *Journal of Politics* 72 4 2010. 957-975.

Giovannini, Enrico, Michela Nardo, Michaela Saisana, Andrea Saltelli, Stefano Tarantola, and Anders Hoffman. *Handbook on Constructing Composite Indicators. Methodology and User Guide*. OECD, OECD Statistics Working Paper STD/DOC(2005)3, Paris 2005, doi:10.1787/533411815016.

Goertz, Gary. *Social Science Concepts: a User's Guide*. Princeton 2006.

Hall, Peter, and David Soskice (ed.). *Varieties of Capitalism. The Institutional Foundation of Comparative Advantage*. Oxford 2001.

Howell, Chris. "Varieties of Capitalism. And Then There Was One?" *Comparative Politics* (36) 1 2003. 103–124.

Knack, Steve, Mark Kugler, and Nick Manning. "Second Generation Governance Indicators". *International Review of Administrative Sciences* (69) 3 2003. 345–364.

Laakso, Markku, and Rein Taagepera. "Effective Number of Parties. A Measure with Application to Western Europe". *Comparative Political Studies* (12) 1 1979. 3–27.

Matthes, Jürgen, and Christoph Schröder. "Rahmenbedingungen für Unternehmen - Zur Aggregation von Weltbankdaten". *IW-Trends* 4 2004. 1–20.

Merkel, Wolfgang. "Embedded and Defective Democracies". *Democratization* (11) 5 2004. 33–58.

Munck, Gerardo L., and Jay Verkuilen. "Conceptualizing and Measuring Democracy. Evaluating Alternative Indices". *Comparative Political Studies* (35) 1 2002. 5–34.

North, Douglass C. *Institutions, Institutional Change and Economic Performance*. Cambridge 1990.

Nunberg, Barbara. *Ready for Europe: Public Administration Reform and European Union Accession in Central and Eastern Europe*. Washington D.C. 2000.

OECD. *Modernising Government: The Way Forward*. Organisation for Economic Cooperation and Development, Paris 2005, http://www.oecd.org/document/15/0,2340,en_2649_33735_35405455_1_1_1_1,00.html.

Pierre, Jon, and B. Guy Peters. *Governing Complex Societies. Trajectories and Scenarios*. New York 2005.

Radaelli, Claudio. "The Europeanization of Public Policy". *The Politics of Europeanization*, Ed. Kevin Featherstone and Claudio Radaelli, Oxford 2003. 27–56.

Sartori, Giovanni. *Parties and Party Systems. A Framework for Analysis*. Cambridge 1976.

Schedler, Andreas. "Conceptualizing Accountability". *The Self-Restraining State: Power and Accountability in New Democracies*, Ed. Andreas Schedler, Larry Diamond, and Marc F. Plattner, Boulder 1999. 13–52.

Schmitter, Philippe C. "The Ambiguous Virtues of Accountability". *Journal of Democracy* (15) 4 2004. 47–60.

Schnapp, Kai-Uwe, and Phillip Harfst. "Parlamentarische Informations- und Kontrollressourcen in 22 westlichen Demokratien". *Zeitschrift für Parlamentsfragen* (36) 2 2005. 348–370.

Thelen, Kathleen, and Sven Steinmo. "Historical Institutionalism in Comparative Politics". *Structuring Politics. Historical Institutionalism in Comparative Politics*, Ed. Sven Steinmo, Kathleen Thelen, and Frank Longstreth, Cambridge 1992. 1–32.

Tsebelis, George. *Veto Players. How Political Institutions Work*. New York/Princeton 2002.

Weaver, R. Kent, and Bert A. Rockman eds. Do Institutions Matter? Government Capabilities in the United States and Abroad. Washington, D.C 1993.

Wiesenthal, Helmut. "Konventionelles und unkonventionelles Organisationslernen. Literaturreport und Ergänzungsvorschlag". *Zeitschrift für Soziologie* (24) 2 1995. 137–155.

Williamson, John, and Stephen Haggard. "The Political Conditions for Economic Reform". *The Political Economy of Policy Reform*, Ed. John Williamson, Washington D.C. 1994. 527–596.

Appendix:

Status Index: Different standardization methods compared

	[1,10]-standardization		z-standardization		log-standardization		Rank differences	
Country	**Score**	**Rank**	**Score**	**Rank**	**Score**	**Rank**	**[1,10] – z**	**[1,10] – log**
Australia	7.75	9	7.51	9	7.40	9	0	0
Austria	6.86	17	6.63	17	6.55	18	0	-1
Belgium	7.17	16	6.94	16	6.92	15	0	1
Canada	7.89	7	7.64	7	7.53	7	0	0
Chile	6.51	22	6.25	22	6.30	22	0	0
CzechRepublic	6.78	18	6.56	18	6.60	17	0	1
Denmark	8.34	5	8.07	5	7.94	5	0	0
Finland	8.52	3	8.24	3	8.12	3	0	0
France	6.74	19	6.49	19	6.47	19	0	0
Germany	7.77	8	7.55	8	7.48	8	0	0
Greece	5.12	30	4.97	30	5.08	30	0	0
Hungary	5.94	25	5.74	25	5.82	25	0	0
Iceland	7.65	10	7.36	11	7.23	12	-1	-2
Ireland	7.37	14	7.15	14	7.07	14	0	0
Italy	5.70	27	5.52	27	5.65	26	0	1
Japan	6.54	21	6.34	20	6.29	23	1	-2
Luxembourg	7.60	12	7.33	12	7.25	11	0	1
Mexico	5.25	29	5.28	29	5.29	29	0	0
Netherlands	7.63	11	7.37	10	7.30	10	1	1
NewZealand	8.51	4	8.21	4	8.06	4	0	0
Norway	8.64	2	8.36	1	8.23	1	1	1
Poland	6.33	24	6.23	23	6.34	21	1	3
Portugal	6.59	20	6.33	21	6.41	20	-1	0
Slovakia	5.48	28	5.33	28	5.39	28	0	0
SouthKorea	5.82	26	5.62	26	5.65	26	0	0
Spain	6.35	23	6.09	24	6.15	24	-1	-1
Sweden	8.65	1	8.36	1	8.23	1	0	0
Switzerland	8.12	6	7.87	6	7.73	6	0	0

Turkey	4.65	31	4.64	31	4.74	31	0	0
UnitedKingdom	7.22	15	6.99	15	6.91	16	0	-1
UnitedStates	7.49	13	7.27	13	7.23	12	0	1

Z-standardization: Quantitative indicators (x) are standardized by subtracting the item-specific mean ($\bar{x}$) and dividing by the standard deviation.

$$y = \frac{x - \bar{x}}{SDEV_x}$$

Standardization based upon a logistic function: This procedure reduces the influence of extreme values while preserving the distance information of indicators. In a first step, quantitative indicators are z-transformed. The standardized values (x) are included into the following logistic function:

$$F(x) = 1 + \frac{9}{1 + e^{-c \cdot x}}$$

The higher c, the steeper F(x) and the more small differences from the mean are expanded. Following Matthes and Schröder (2004), we calculate c as the square root of the coefficient of variation of the respective indicator. This method ensures that equal absolute differences are not dependent on the value of the mean and have a larger effect if the variation is low.

Management Index: Different standardization methods compared

	[1,10]-standardization		**z-standardization**		**log-standardization**		**Rank differences**	
Country	**Score**	**Rank**	**Score**	**Rank**	**Score**	**Rank**	**[1,10] – z**	**[1,10] – log**
Australia	7.71	6	7.56	6	7.46	6	0	0
Austria	6.39	16	6.37	15	6.28	17	1	-1
Belgium	6.00	21	5.89	23	5.83	26	-2	-5
Canada	7.04	10	7.02	9	7.13	8	1	2
Chile	6.15	18	6.11	18	6.01	21	0	-3
CzechRepublic	5.88	23	5.83	25	5.91	25	-2	-2
Denmark	7.90	3	7.87	3	7.72	4	0	-1
Finland	7.79	4	7.68	5	7.76	3	-1	1
France	5.82	25	5.86	24	5.97	22	1	3
Germany	6.84	11	6.81	14	6.72	13	-3	-2
Greece	4.54	31	4.54	31	4.51	31	0	0
Hungary	5.71	28	5.65	28	5.72	28	0	0
Iceland	7.23	8	7.21	7	7.13	8	1	0

Ireland	6.33	17	6.25	17	6.33	16	0	1
Italy	5.62	29	5.61	29	5.52	29	0	0
Japan	6.41	15	6.27	16	6.34	15	-1	0
Luxembourg	7.05	9	7.00	10	6.88	12	-1	-3
Mexico	5.87	24	5.92	22	6.03	20	2	4
Netherlands	6.84	11	6.83	13	6.72	13	-2	-2
NewZealand	7.72	5	7.72	4	7.62	5	1	0
Norway	8.20	2	8.11	2	8.05	2	0	0
Poland	5.79	26	5.82	26	5.95	23	0	3
Portugal	5.76	27	5.72	27	5.83	26	0	1
Slovakia	4.75	30	4.78	30	4.89	30	0	0
SouthKorea	5.92	22	6.06	20	6.13	18	2	4
Spain	6.03	20	6.07	19	6.04	19	1	1
Sweden	8.29	1	8.27	1	8.15	1	0	0
Switzerland	6.79	14	6.88	11	6.96	10	3	4
Turkey	6.07	19	6.02	21	5.93	24	-2	-5
UnitedKingdom	6.82	13	6.86	12	6.90	11	1	2
UnitedStates	7.24	7	7.03	8	7.15	7	-1	0

This article is part of the Sustainable Governance Indicators 2011 project.

Contact:

Bertelsmann Stiftung
Carl-Bertelsmann-Straße 256
33311 Gütersloh

Dr. Daniel Schraad-Tischler
daniel.schraad-tischler@bertelsmann-stiftung.de

Najim Azahaf
najim.azahaf@bertelsmann-stiftung.de