EVALUATING EXECUTIVE GOVERNANCE IN OECD MEMBER STATES. DESIGN AND KEY FINDINGS OF AN EXPERT SURVEY

Paper for the ECPR General Conference,
Section: Policy Evaluation in the Public Sector

University of Potsdam, 10-12 September 2009

DRAFT VERSION

Martin Brusis, University of Munich, martin.brusis@lrz.uni-muenchen.de
Anna Naab, University of Munich, naab@lrz.uni-muenchen.de
Jürgen Turek, University of Munich, juergen.turek@lrz.uni-muenchen.de

ABSTRACT

This paper proposes to conceptualize governance as a synthesis of the governing capacity and political accountability of core executives. The concept of *executive capacity* focuses on the institutional capabilities enabling core executives to take strategic decisions. Key functions and processes include the preparation and formulation of policies, the implementation of policies, the incorporation of foreign reform impulses and the learning of (or in) institutions. The concept of *executive accountability* studies how actors in the wider political system and society—citizens, parliaments, parties, interest associations and media—are able to hold the core executive accountable, provide knowledge and deliberate the norms guiding its decisions.

These concepts have been operationalized in questions and quantitative indicators that were assessed and compiled by country experts for 30 OECD countries in 2007. A key finding of this survey has been that higher levels of executive accountability support executive capacity. Countries with particularly high capacity and accountability have established superior arrangements for institutional learning. Countries with weaker executive governance are often characterized by dominant executives and relatively weak non-governmental actors.

Introduction

The global financial and economic crisis has required and legitimized extraordinary governmental intervention in times that are dominated by skepticism about the capacity of governments to take prudent, longterm-oriented and collectively rational decisions. Numerous trends indicate an erosion of governments' capacity to govern: The economic crisis places enormous fiscal burdens on states whereas their revenue-raising powers are structurally limited by international tax competition. Interdependencies between policies have increased both sectorally and cross-nationally in the wake of economic globalization and transnational integration. National political systems are prone to partisan political blockades that may only be overcome through package deals that often fail to address the real problems. Imperatives of media attention dictate the political agenda, rendering strategic deliberation and problem-solving secondary. Governments have responded to blockades and interdependencies by transferring decisionmaking to agencies and independent expert bodies which has exacerbated problems of democratic accountability. Opinion surveys document a decline of trust in political institutions.

While governments today seem to be less able to provide good governance, their role is difficult to avoid or to replace. Possible alternative governance providers suffer from various restrictions and shortcomings: International organisations are constrained by unanimity requirements; parliaments tend to be sidelined in internationalized decisionmaking processes that are tailored to national executives; parties face an erosion of stable electoral milieux and the stable, programmatic membership parties of the past seem to be turn into electoral parties; civil society organisations possess only limited capacities and their democratic legitimacy is at times questionable; finally, companies increasingly depend on international capital markets that tend to favor short-term rentability concerns over longterm strategic interests.

The erosion and simultaneous need for governmental leadership have created both advocacy and scholarly motives for the "Sustainable Governance Indicators" (SGI), an expert survey and dataset we present in this paper. The SGI have been developed by two thinks tanks based in Germany, the Bertelsmann Foundation and the Center for Applied Policy Research at the University of Munich. Our general aim has been to investigate and measure the reform capacity of states and to raise the awareness of the public and politicians for the quality of governance.

The present paper explains how this interest in reform capacity—the capacity of political actors to identify and implement changes needed to improve the status quo—has led us to an indirect approach of measurement that assesses both the policy performance of states and the institutional arrangements of governing, termed "executive governance". We conceive executive governance as the interaction of core executives with an enabling environment of political actors. We assume that the institutional potential for reform can be considered high if core executives have the capacity to take strategic decisions and if citizens, parliaments, parties, interest associations and media are able to hold core executives accountable, provide societal knowledge and deliberate the norms guiding these decisions. In the following sections of this paper we elaborate these concepts and how we have attempted to measure them. We then report key findings of the survey, with a focus on the quality and patterns of executive governance.

1. Measuring reform capacity: a review of the literature

Three approaches of comparing the reform capacity of states or governments may be distinguished.

(1) A first approach is to establish a catalogue of concrete reform measures and to examine the extent to which countries have realized these reforms. The more reforms a state has implemented, the higher its capacity of reforming. Examples of such clearly delineated reform programs are the "Washington Consensus" of the International Monetary Fund and the World Bank, or the "New Public Management" (Pollitt/Bouckaert 2004; Williamson 1994).

Yet one can doubt whether the implementation of given reform catalogues allows inferences on reform capacity in the sense of a structurally engrained potential. The core meaning of reform capacity might rather be the dynamic adaptability to new challenges which is not captured by studies that assess the extent to which predefined reform measures have been implemented. Which policies and institutions are more adaptable could thus be measured only indirectly and retrospectively, by comparing policy outputs and/or outcomes at two subsequent points of time (Wiesenthal 2005). Improvements over time would allow to infer a reform capacity. Such an evidence-based measurement of reform capacity would also avoid the political bias often associated with specific policy measures and the beliefs underpinning their application. This is why evidence-based comparisons of policy performance are increasingly accepted in international cooperation frameworks, as shown by the United Nations' Millenium Development Goals or the EU's Lisbon Process.

(2) This second approach is inter alia advocated by economists who assess the efficiency of public expenditures by comparing the levels of expenditure in 23 OECD states, new EU-member states and "emerging markets" (Afonso et al. 2005; 2006). Afonso, Schuknecht and Tanzi conceptualize public sector performance as comprising allocation, distribution and stabilization functions as well as the provision of equal opportunities. To evaluate performance, they use quantitative indicators intended to describe the quality of public administration, education, health, transport and communication infrastructure (equal opportunities) as well as income distribution, growth, inflation and unemployment. These indicators are aggregated by calculating means from the respective equally weighted standardized components. The ratio of public sector performance and public expenditures (measured as a share of GDP) indicates the efficiency of the public sector, which according to the authors is higher in countries with lower public expenditure shares.

In contrast with this simple input-output-comparison, other studies try to identify different profiles of public sector performance and explain them with institutional determinants. The Social and Cultural Planning Office (SCP), a Dutch governmental think tank, calculated aggregate performance scores for the public sector from 19 quantitative indicators collected for 22 developed industrial states (2004). The indicators are divided into four groups: stabilization and economic development; distribution (measured by poverty rates); allocation of public services; quality of the public administration. The composite measure is determined by calculating first the group means of each individual, z-transformed and equally weighted indicator and second an overall mean of the equally weighted group means.[1]

Roller measures policy outcomes in the fields of internal security, economic policy, social policy and environmental policy by collecting 14 indicators for 21 OECD countries over the period from 1974 to 1995 (Roller 2006). She constructs a composite indicator of "political effectiveness" similar to the SCP study, in a two-step mean-based aggregation procedure that

[1] Unlike Afonso et al., the SCP study defines the volume of public expenditure based on a functional definition of the public sector, which also includes the private production of services that are traditionally considered as public services.

assigns equal weights to policy areas and to individual indicators which have been standardized through a linear transformation.

Yet both SCP and Roller can explain the influences of institutional factors on policy outcomes only partially and their lessons for institutional reform remain vague. In the tradition of Esping-Anderson (1990), SCP classifies states by structural features of public services (use of resources, funding, delivery). However, the geographical-cultural groups of Anglosaxon, northern, western, southern and central European states resulting from this classification suggest little scope for intentional model change.

Roller also distinguishes patterns of political performance, but drawing on Lijphart (1999) she additionally classifies states according to the extent of power dispersion, which is determined by characteristics of the constitution and structures of government and opposition. But she finds only selective and inconsistent effects of negotiation / majoritarian democracy on the level, stability and structure of political effectiveness (Roller 2006, 276-277). She interprets this result with the fact that her classification does not sufficiently take the breadth of interest representation into account. However, one can also ask more fundamentally whether her decision to operationalize power dispersion through veto player indices has been appropriate. Considering only few macro-institutional features turns gradual into categorical differences and sacrifices minor differences between institutional arrangements to the theoretical interest in parsimony. Furthermore, macro-institutional classifications of presidential or parliamentary, bi- or unicameral, federal or unitarian systems do not offer viable starting points for institutional reforms, since such fundamental features of political systems are usually beyond the discretion of political actors.

(3) These questions and objections lead to a third analytic and conceptual approach that conceives reform capacity as a function of the political process and its management. At levels below the macro-institutional categorizations, an increasingly accepted and expanding body of desirable public management practices has evolved among practitioners. International organizations and agencies like the World Bank, OECD and EU Commission use these agreed good practices as benchmarks for international comparisons (Evans/Manning 2003; James/Ben-Gera 2004; Manning 2005; Nunberg 2000; OECD 2005). These organizations try to develop process-oriented standards because they consider good governance a crucial factor of successful development and view benchmarks as effective orientations and incentives for domestic reforms in the framework of performance and new public management approaches (Knack, Kugler and Manning 2003; Kuhlmann, Bogumil and Wollmann 2004).

With the Governance Indicators, World Bank economists have developed a sophisticated methodology of governance measurement that synthesizes indicators of governance quality in 213 countries and territories from 25 different organizations and 31 data sources (Kaufmann et al. 2006). Governance is evaluated by six different composite indicators that refer to the accountability of governments, their implementation capacity and the acceptance of political institutions. Of these indicators, “government effectiveness” seems to best represent the contribution of governments to reform capacity, since it describes “the quality of public service provision, the quality of the bureaucracy, the competence and independence of the civil service, and the credibility of the government’s commitment to policies. The focus is on inputs the government needs to produce and implement good policies and deliver public goods.” (Kaufmann u.a. 2004: 255). The data sources for this indicator are expert polls, business and popular opinion surveys carried out by the World Bank as well as by non-governmental organizations, universities and commercial rating agencies.

From these sources, items associated with government effectiveness are selected, standardized and weighted according to their representativity and precision. This procedure allows to estimate governance as “the mean of the distribution of unobserved governance conditional on the […] observed data points" for a country (Kaufmann, Kraay, and Mastruzzi 2004, 259). This aggregation and estimation procedure tries to filter out the valid information content of the data sources by assigning lower weights to sources that differ strongly from the assessment of other sources. Critics have pointed out that this procedure weights similarly biased sources higher due to their correlation, which leads to less precise governance estimates (Arndt/Oman 2006). An even more serious problem is the loss of conceptual precision caused by the aggregation procedure that reduces items from different sources varying in their wording and meaning to numeric information that is then reconstructed in a new indicator (Knack 2007; Van de Walle 2006).

Contrary to the World Bank's synthetical Governance Indicators, the OECD disaggregates government activities into a wealth of cross-nationally comparative qualitative and quantitative indicators (OECD 2007). Based on the model of a production process, the OECD in early 2007 published comparative data on government revenues, inputs, processes and outputs. With 32 individual data series, the process data constitute a focus of the compilation and include public budgeting, human resource management, monitoring, regulatory quality, the role of centers of government and e-government. However, most datasets have been selected from already available, already published OECD survey and fiscal data. Moreover, the compilation does not explicate whether the datasets are supposed to cover the respective part of the governmental production process in an exemplary fashion (as proxies) or whether they provide only fragmentary information.

Until now, the OECD has not constructed any composite indicators from those dingle data and intends to proceed carefully since such indicators would suggest „a spurious degree of precision in inter-country ranking" (OECD 2007, 7). It is remarkable that this critique addresses only the risks of misinterpretation, but does not raise principled doubts regarding the aggregability of institutional and process features. The OECD does consider the data a suitable basis for measuring efficiency and effectiveness (OECD 2007, 2) and thus for a graduating international comparison. Most process-related questions and response options imply an obvious interpretation regarding good and less good administrative practice and thus at least an ordinal level of measurement.

2. Design of the Sustainable Governance Indicators

The SGI do not venture on a catalogue-based measurement of reform capacity. Rather, the conceptual design of the survey seeks to link and develop approaches of results- and process-oriented measurement. Reform capacity is assessed indirectly by comparing the policy performance and executive governance of states. A cross-national comparison of policy outcomes allows reform activities and their impact to be assessed *ex post*. Better policy outcomes or greater improvements – cross-nationally or cross-temporally – may result from a higher reform capacity in a given country. As the international consensus on "good" policy outcomes grows, identifying reform capacity *ex post* has become much more feasible.

In addition, the SGI adopt an *ex ante* perspective and try to assess the institutional capacity for reforms by comparing "executive governance". This concept denotes the institutional arrangements of governing, including the mechanisms for and patterns of interaction between the core executive and its organizational environment, both within the executive itself and in the wider political system. Executives have command over significant resources and translate popular preferences into policies. How they are governed may not guarantee the success of reforms, but it surely affects the chances of governments to succeed with their reform measures.

To evaluate executive governance, the SGI focus on the micro-level functions and processes required to "run a government", drawing on the evolving international public management know-how in areas such as strategic planning, inter-ministerial coordination, the drafting of legal acts, monitoring, budgeting, auditing, task delegation, institutional learning devices, and public communication and consultation policies. We assume that the extent to which a government has established best practices in performing its functions can be taken as an indication of better executive governance that is likely to enhance a country's institutional capacity for reform.

Policy performance and executive governance are represented in two composite indicators: a Status Index and a Management Index. These two indices consist of 149 individual items–93 and 56, respectively. Seventy-four quantitative indicators are derived from information collected from public data sources. Experts for each country have provided 62 qualitative assessments as well as 13 quantitative indicators.[2]

Figure 1: Composition of the Status and Management Indices

	Status Index	Management Index
Dimensions	2	2
Categories	4	7
Criteria	18	15
Items	93	56
Of which: Expert assessments	26	36
Quantitative indicators	67	7
Quantitative expert indicators	-	13

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 Index therefore includes questions on democracy that assess 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. To assess policy performance, the Status Index then examines four broad policy sectors and policy areas:[3]

(1) Economy and employment: labor market policy, enterprise policy, tax policy, budgetary policy;

(2) Social affairs: health policy, social cohesion, family policy, pension policy;

(3) Security: external/internal security policy; integration policy;

(4) Sustainability: environmental policy, research and innovation policy, education policy.

[2] For a detailed list of indicators, questions and sources, please see www.sgi-network.de.
[3] In addition to this policy-specific performance assessment, the Index is also based on indicators that reflect the broad socioeconomic performance of countries (e.g., potential GDP growth).

Each policy area was evaluated by country experts and through indicators from public data sources. Country experts were asked to assess a set of questions and evaluate the extent to which a particular policy realizes specified objectives, such as the goal of fiscal sustainability in the case of budgetary policy. These objectives have been carefully selected and defined so as to avoid any ideological bias and to make sure that they would be broadly accepted and supported by citizens, policymakers and scholars alike across both political and value-based divisions.

In selecting the performance indicators from public data, we have also 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 sought to avoid including model-specific indicators that might be seen as being biased in favor of "coordinated" or liberal market economies (Hall and Soskice 2001; Howell 2003). The key rationale for combining democratic, socioeconomic and policy performance items in one index has been that high democratic and socioeconomic standards constitute necessary scope conditions for policy-specific performance.

The Management Index

The Management Index reflects the consensus practitioners and scholars have developed on what good governmental practices 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 and specifically the core executive as the key actor of governance (Knack, Kugler and Manning 2003). 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 their normative appropriateness. This dimension of "executive accountability" reflects the degree of importance attained in governance by actors outside the executive.[4]

Both dimensions of executive capacity and executive accountability are further structured into categories and criteria (see figure 2). The categories used to assess executive capacities refer to stages in the policy cycle as well as to concepts deriving from the literature on Europeanization, globalization and policy learning (Common 2004; Dolowitz and Marsh 2000; Radaelli 2003):

(1) Policy preparation: strategic planning and expert advice, inter-ministerial coordination, regulatory impact assessment (RIA), consultation and communication policies;

(2) Policy implementation: anticipation of veto actors in the legislative process, management of task delegation to ministers, agencies, subnational governments and private actors;

(3) Incorporation of external reform impulses: governmental capacity to adapt to globalization, Europeanization or transnationalization as well as to import and export policies;

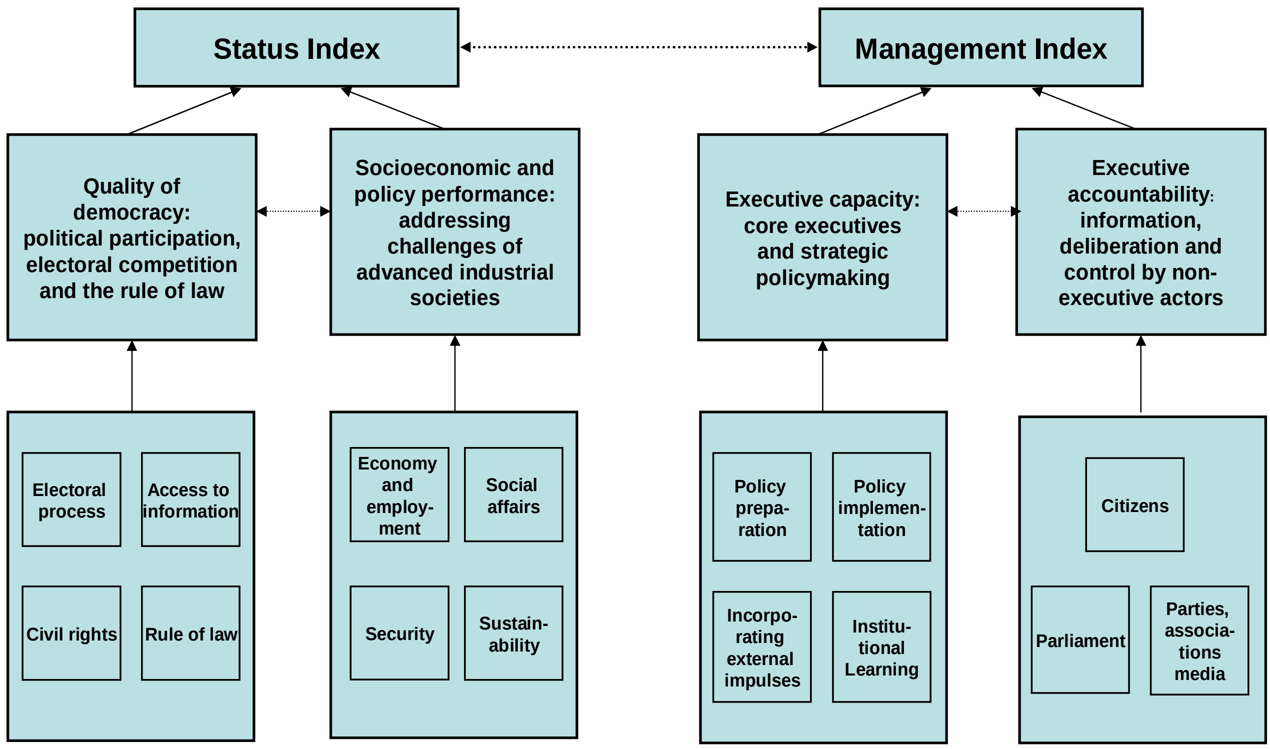
(4) Institutional learning: governmental capacity to reform its own institutional arrangements and improve its strategic orientation.

---

[4] 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)

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 particular questions here ask: To what extent are citizens informed about government policies? Is the parliament capable of evaluating and controlling the executive? And are intermediary organizations (the media, political parties, interest associations) characterized by policy know-how and relevance? These questions were evaluated by country experts who also collected numerical data on, for example, the share of governmental bills adopted by parliament or the size of expert staff in parliament.

Figure 2: Components of the Status and Management Index



In sum, the design of the Management Index assumes that more accountability – in the form of public scrutiny, information channels and normative deliberation – improves executive governance. Since the conventional approach of governing tends to view accountability and participation mechanisms as constraints on executive authority, this assumption might be challenged as counterintuitive. Reforming may be much easier for some governments as they can govern under much more conducive structural conditions and rely on enabling actor constellations.

This critique can be substantiated with the theoretical argument that a greater number of ideologically opposed and internally coherent veto players increases an adopted policy's degree of stability (Tsebelis 2002). However, veto players do not necessarily block improvements to the status quo. In fact, many authors have argued that veto players might even improve the quality of reforms by helping governments to better assess the potential impacts of a given reform. Doing so induces reformers to broaden their bases of support by accommodating veto-player critiques and thereby rendering reforms irreversible (c.f.,Benz 2003, 230).

For this reason, we do not attribute a veto function a priori to certain attributes of a given political system, such as a large number of (governing) parties, a bicameral parliament or a strong constitutional court. Instead, we use empirical data to examine whether veto players have a positive influence on reform policy or whether they just block it, driven by confrontational preferences. In addition, we switch perspectives and ask to what extent governments are able to anticipate veto players in the legislative process (Evans and Manning 2003).

One might still argue that each OECD country has a different number of veto points (i.e., potential blockade constellations) that their respective governments need to consider in legislative processes. For example, whereas British governments do not have to take into consideration either a constitutional court's concerns or those of a second parliamentary chamber with veto rights, German governments must take these two formidable veto players into account.

This comparison naturally raises the question as to whether a German government that must anticipate more veto points in the lawmaking process should have its success more positively evaluated than a British government that, in comparison, has fewer veto points requiring consideration.

For the purposes of the SGI, we have decided to treat states equally in this respect, regardless of whether their governments must anticipate two, three, four or five veto points. One could object to this methodology by pointing to the aforementioned example and argue that it is easier for the British government to attain the best assessment because it effectively has to anticipate only two veto points. At the same time, however, governments in systems with more veto points than the United Kingdom are aware of the additional veto points and therefore can—and indeed must—prepare for them accordingly. To permit de facto a bonus for states with veto-intensive systems could result in a situation in which such a state—despite having had several proposed laws fail approval by the constitutional court—is accorded a better score than a government in a system with less veto points that succeeded in having more laws adopted.

Furthermore, according to this methodology, a system based on the majority principle (e.g., the United Kingdom's) does not automatically receive a good rating for strategic capacity just for having comparatively low veto hurdles. Instead, a country's aggregate assessment also takes into account the assessments of the government's consultations with business and social actors and its communication with the public. For example, a British government that, thanks to its majority principle, is able to pass many laws while, at the same time, ignore societal interests, would receive a poor rating for the respective question. 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.

3. Measurement and aggregation

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 by subjective perceptions and thereby pose problems of intercoder reliability (Munck and Verkuilen 2002).The SGI's expert survey questionnaire was designed to improve the validity of expert assessments through the use of six tools and procedural steps.

(1) Many assessment questions are formulated so as to elicit detailed factual evidence rather than broad—and consequently more subjective—assessments.

(2) The questionnaire provides detailed explanations of and four tailored response options for each question. The experts were instructed to adapt the standardized response options to the individual context of the particular country they were 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.

(3) Each OECD member state was examined by three leading scholars with established expertise in their 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. Each expert was tasked with writing assessments for "their" country, which resulted in the production of three individual and yet parallel country assessments ("expert reports") for each country. The experts were instructed to assess the situation in their countries as of March 2007 and to take into account the period between January 2005 and March 2007 when explaining their ratings.

In completing the questionnaire, each expert had to provide ratings for 62 questions, which means that the evaluations for all 30 countries entailed a total of 1,860 scores, or ratings. For 65 percent of the ratings, the expert scores deviated by two levels or fewer, which results in a standard deviation of equal to or less than 0.94. However, these deviations tend to overestimate the imprecision of the ratings, as the motivation behind selecting three 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 validity 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 reviewers, 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.

(4) The countries examined by the SGI were subdivided among seven regional coordinators according to geographical location. These regional coordinators, who are political scientists with both comparativist and area expertise, were each responsible for three to five of the 30 OECD countries. Selecting information from each expert report according to the criteria of validity and objectivity, the regional coordinators integrated this information into a synthesized "country report" and gave numerical ratings based on those provided in the three expert reports.

(5) The regional coordinators reviewed their ratings collectively so as to make it possible to draw comparisons across the entire OECD world. As part of the discussions making up the

review process, each regional coordinator was required to explain, defend and eventually recalibrate his ratings and assessments. To make any changes agreed to during the review process more transparent, the coordinators also agreed to choose the median of the three country expert ratings as the default score and, if a deviation from the median score was deemed necessary, to keep the score within the range of ratings provided in the experts' reports. During the review process, the regional coordinators deviated from the respective median values in 31 percent of the total number of 1860 scores provided in the reports. In turn, two 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.

(6) As part of a second round of reviews, an advisory body composed of renowned scholars and practitioners and charged with making strategic decisions discussed and approved the ratings. This second review resulted in changes to 6 percent of the total scores and entailed a slight reduction in the proportion of scores deviating from the country expert median in the expert reports (down to 29 percent from 31 percent) and exceeding the range of the ratings in the experts' reports (down to 1 percent from 2 percent).

The SGI's other main sources of data are quantitative indicators collected from publicly available statistics. 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 had to first 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.

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 (OECD/JRC 2008). 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 (Brusis 2009). These standardization methods produced very similar rankings, with the scope of rank shifts being limited to a maximum of three ranks.

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 3 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.

This benchmarking technique made full use of the range given by the scale. At the same time, however, it also caused a divergence between indicators and expert ratings, as the latter rarely chose the lowest possible scores. This means that the choice of empirical, relative benchmarks led to a situation in which, for the quantitative indicators, the worst performers were assigned a score of one even if they performed only slightly worse than other countries. In contrast, without a method of standardization for the expert ratings, the worst performers here would have "suffered" less from a small gap in relation to better-performing countries. Since we did not want to treat quantitative indicators and expert ratings in different ways, we rescaled all expert ratings so as to generate distributions with identical ranges. To avoid the emergence of any discrepancies between expert ratings and written assessments, we also included the original, non-standardized expert ratings in the country reports.

In order to integrate individual items into a composite index, weights have to be assigned to all individual items. The SGI's method of weighting these items has been guided by two main 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. In contrast, our prior empirical knowledge about, for example, the impact of effective inter-ministerial 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, for example, on how inter-ministerial 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 inter-ministerial coordination as a component of policy preparation, one must assume that effective inter-ministerial 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 how much 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 has 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.

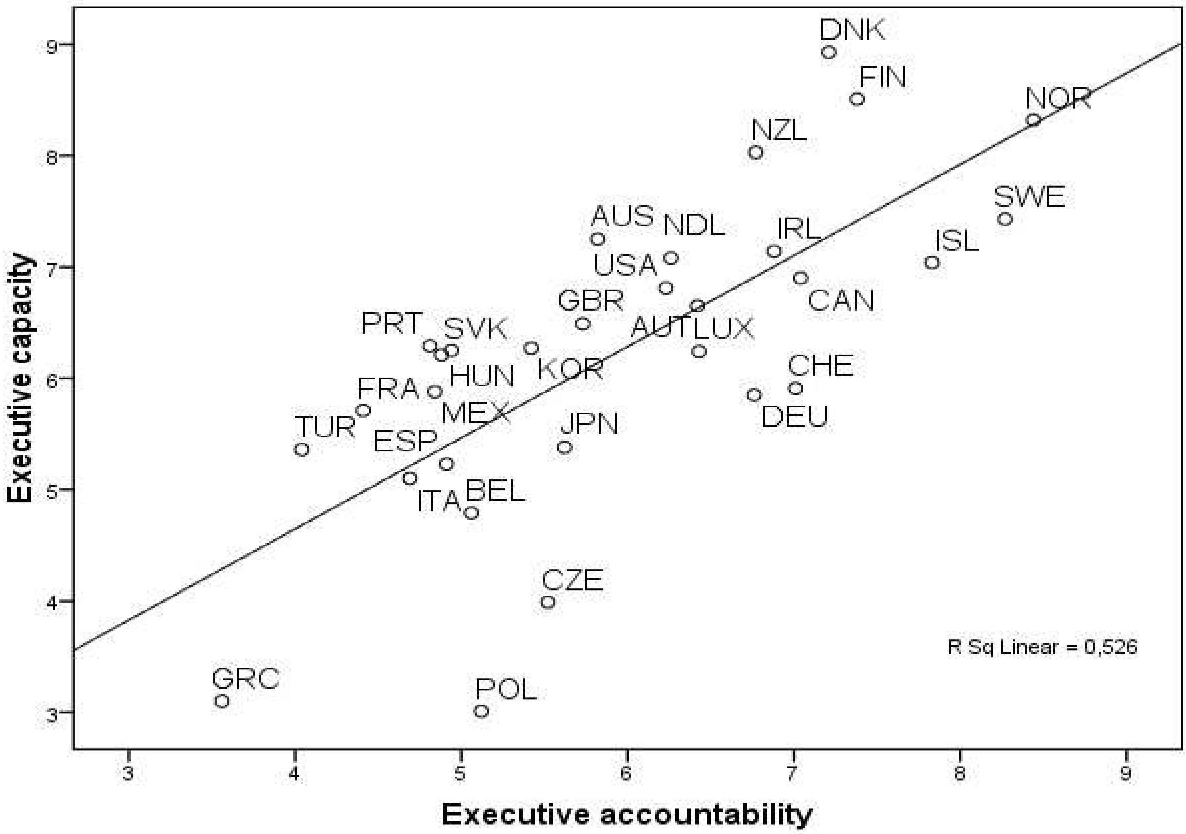
In this way, the combination of expert assessments and statistical indicators assumes that both types of observations have specific strengths and weaknesses, that they cannot fully substitute for each other, and that neither of them is epistemologically superior to the other (Collier, Brady and Seawright 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.

The Status Index and the Management Index scores are thus derived by calculating the arithmetic means of the scores for their respective two components (i.e., dimensions, categories and criteria).[5] In order to confirm the robustness of our chosen approach of weighting, we tested various other weighting models. The aggregate correlations between these models and our default model turned out to be fairly high, which confirms the robustness of our weighting.

4. Key findings

This section summarizes key findings of the survey, focused on the Management Index. We start by describing and interpreting the grouping of countries that emerges from the ranking. Then we analyze patterns of centralized vs. segmented executives, policy coordination and the management of delegation in federal states. A first, aggregate-level finding is that the two dimensions of the Management Index—"executive capacity" and "executive accountability"— are highly correlated. This correlation confirms our assumption that accountability does not have to constrain a government's decisionmaking capacity and may even increase executive capacity.

Figure 3: Management Index results



[5] As 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 were not used as levels in the aggregation process.

Patterns of executive governance

Four different state groupings emerge from a ranking of the 30 countries based on their aggregate Management Index scores. Norway, Denmark, Finland, Sweden, Iceland and New Zealand hold the top positions. This leading group sets itself apart from the lower-ranked states particularly with respect to its institutional learning capacity. In all of the countries within this group, the various parts of the executive are continually scrutinized and frequently reformed so as to improve their efficiency and performance capacity. Most reforms are guided by strategic considerations rather than by partisan political interests, and are primarily concerned with restructuring the portfolios of the various ministries, as well as relations between ministries and semiautonomous agencies.

On the other hand, the differences between the leading group and the intermediate group are less pronounced with respect to policy preparation and formulation, parliamentary monitoring and information-gathering resources or intermediate organizations are involved. In these areas, Iceland and New Zealand performed more poorly than many other states from the upper part of the intermediate group.

The upper intermediate group of the Management Index is made up of states including Ireland, Canada, Australia and the United States, along with the smaller Central European states of Austria, Switzerland, Luxembourg, the Netherlands and, at the end of this group, Germany. These states display a weaker institutional learning capacity than those in the leading group, but also have responded in a less active and innovative way to globalization and Europeanization, and they have a smaller proportion of well-informed citizens. They differ from the countries in the lower intermediate group particularly in terms of their party, media and association systems, all of which tend to provide more support to governing.

The aggregate scores of the individual dimensions confirm the hypothesis that, in the countries belonging to the Anglo-American traditions with majoritarian democracies, executive capacities are stronger while the monitoring and information-gathering functions of societal actors are weaker. In the Central European countries, the relationship is exactly the opposite: Governance is supported less by the executive’s institutional and strategic capacity and more by the monitoring, information-gathering and deliberative capacities of parliaments, parties, media organizations and interest associations, and by the political knowledge of the citizenry.

The Management Index’s lower intermediate group includes the United Kingdom, South Korea, Slovakia, Hungary, Portugal, Japan, Mexico, Spain, France, Belgium, Italy, the Czech Republic and Turkey. In comparison with the countries in the upper intermediate group, the 13 countries in this group displayed weaknesses particularly in the fields of executive accountability to citizens and parties, as well as in the political expertise and relevance of interest associations. One of the key explanations for this particular shortcoming was mentioned in several country reports: Interest associations in the countries of Southern Europe, but also in Asia and Mexico, attempt to influence government policies largely through networks of political decision makers, and less so by means of public debate contrasting government policies with alternative policy proposals. However, the country reports also pointed out a lack of capacities, and problems with setting priorities.

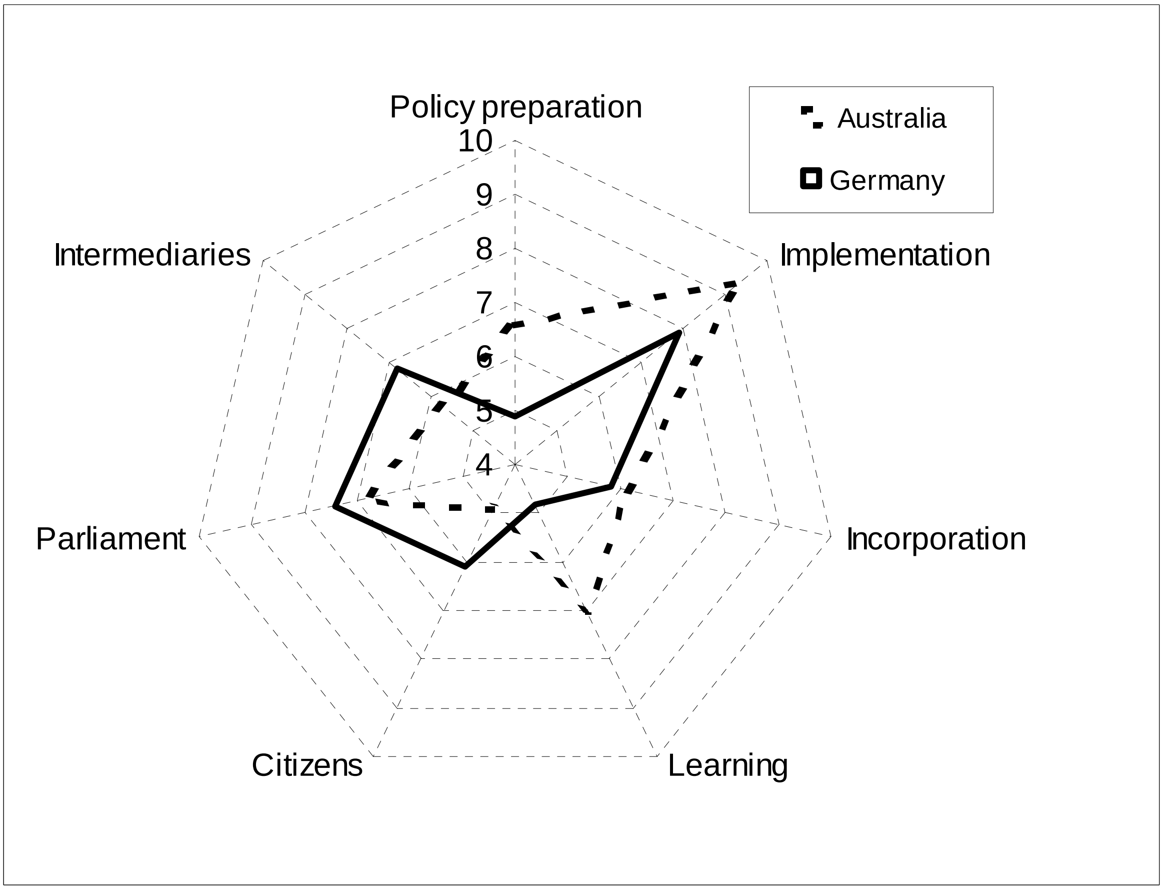
The majority of states in the lower intermediary group have similar models of governance, at least in the sense that executive capacity is clearly dominant as compared to the accountability of the executive to external actors. Six of the 13 states had capacity scores that were even higher than that of Germany, the lowest-scoring country in the upper intermediate group. On the other hand, in terms of executive accountability, none of the 13 countries attained a score as high as those obtained by countries in the upper intermediate group.

Poland and Greece trail the other OECD states to a significant degree. Their deficiencies compared with the intermediate group are particularly pronounced in the field of strategic capacity, while the parliaments of both countries enjoy much better monitoring and information-gathering resources than several of the countries receiving higher scores. In terms of globalization, Europeanization, policy preparation and institutional learning, the governments of Poland and Greece scored significantly lower than the average country in the lower intermediate group.

Centralized vs. segmented executives

The two dimensions of the Management Index allow styles and models of governance to be roughly distinguished as to whether the strategic capacity of the executive takes prominence in decision-making processes or whether the control, information-gathering and deliberation functions of societal actors outside the executive are more strongly pronounced. The differing capacity profiles of Australia and Germany illustrate this distinction (see figure 4). In the Management Index, Australia lies four places above Germany and owes its position in particular to its government's strategic capacity. On the other hand, executive accountability to citizens, the parliament and intermediate organizations received higher scores in Germany than in Australia.

Figure 4: Executive vs. non-executive led governance: Australia and Germany



Delegation management in federal states

Among the nine states in the upper intermediate group, there are six federal states in which most policies are also formulated and implemented in relation to states or provinces, reflecting the political conditions of a multi-level system. The Management Index examines and assesses the extent to which governments effectively delegate tasks to self-administering and self-governing bodies at both the regional and local level. In this process, the degree of decentralization is not necessarily positive (or negative) in itself. What is much more important is whether the central government has created structures which fully exploit the potential of decentralization (more proximity to citizens, more efficient public services and more effective public accountability of office holders). The Index ascertains the extent to which central governments: provide sources of revenue to the regions sufficient to fulfill the tasks that have been delegated to them; respect the constitutionally defined discretion of regions; and ensure regions meet national standards for the delivery of public services.

The assessment profiles associated with these three criteria highlight the different structures of intergovernmental relations in the federal states examined here. Germany and Australia earn higher scores when it comes to safeguarding national standards than for guaranteeing autonomy related to finances and decision-making. This assessment is consistent with the model of cooperative federalism and points to the fact that Germany’s federal government puts a much higher priority on uniformity of living conditions than on safeguarding federal states’ autonomy in legal and financial matters. Australia’s system of vertical financial equalization decouples the responsibilities for expenditures and revenues. The country’s federal government is increasingly taking advantage of its constitutionally guaranteed authority with regards to revenue in order to implement policies that would traditionally fall under the jurisdiction of the individual states and territories. It has tied federal budget allocations to the condition that state governments implement measures set at the federal level.

On the other hand, Switzerland and the United States serve as examples of competitive federalism. In comparative terms, both countries place less importance on safeguarding national standards than on state governments’ autonomy as related to finances and decision-making. Switzerland’s cantons and U.S. states are able to make autonomous decisions about types and rates of taxes as well as about important expenditures. The federal governments monitor compliance with national standards only in individual policy areas and to a limited extent. Nevertheless, in real terms, the Swiss cantons and U.S. states enjoy a lower degree of political, legislative and administrative autonomy than do Germany’s federal states.

Figure 5: Profiles of policy delegation in federal states

Central government provides…	...fiscal-autonomy	...political-autonomy	...country-wide standards
BEL	6	10	7
DEU	7	7	9
ESP	6	7	4
CAN	6	8	7
USA	9	10	6
CHE	10	10	7
AUS	6	5	7
AUT	7	7	6
MEX	7	6	4

Inter-ministerial coordination

In examining the scores related to the role of the center of government, two separate groups of countries can be identified (see figure 6). Italy, Slovakia, Czech Republic and Poland have centers of government that play a subordinate, assisting role in the preparation of cabinet proposals as well as in the monitoring and control of the line ministries. On the other hand, Belgium, Mexico, South Korea, Hungary and the United States have centers of government that are outfitted with their own policy expertise and with far-reaching rights to intervene, and as a result are able to shape governing in a significantly stronger way.

All of the countries of the lower intermediate group that have been examined here as examples have developed what are primarily political mechanisms of interministerial coordination, in which ministers, the head of the government and political leaders reach agreements, set priorities and negotiate compromises on governmental policies. On the other hand, the administrative coordination between specialized civil servants within the ministries is accorded only minor importance. In Belgium, Slovakia, the Czech Republic and Italy, interministerial coordination takes place across the entire cabinet as well as in the context of cabinet committees or informal ministerial meetings, all of which serve as venues for collective decision making. In the remaining countries, one can observe more strictly hierarchical forms of political coordination that are dominated by prime ministers, or in the case of the presidential systems of Mexico and the United States, by presidents.

The relatively large importance given to mechanisms of political coordination differentiates the countries in the lower intermediate group from those in the upper intermediate group and the leading group. Only in the latter two groups were there countries in which the center of government has control by virtue of its expertise and rights of intervention and in which the administrative mechanisms of interministerial coordination are dominant.

Figure 6: Arrangements of interministerial coordination

	Political coordination	**Administrative coordination**
Steering center of government	BEL, MEX, KOR, HUN, USA	IRL, ISL, CAN, LUX, NOR
Assisting center of government	ITA, SVK, CZE, POL	

Classification: Mean of the scores for M3.1, M3.2, M3.3, M9.2b > 8: steering COG, < 6: assisting COG. The classification of political and administrative coordination is based on the scores for M3.4-M3.6 in connection with an analysis of the country reports.

Conclusion

The Sustainable Governance Indicators have been designed to assess reform capacity indirectly by measuring both policy outcomes and executive governance. This indirect approach entails a degree of methodological, theoretical and political self-restraint.

Firstly, we do not assume that greater institutional capacity will necessarily generate improved socioeconomic performance or better performance in terms of policy outcomes or the quality of democracy. These outcomes are also and sometimes decisively influenced by other determinants, such as a favorable international environment, charismatic leadership or fortunate coincidences. However, greater institutional capacity does improve the chances that political leaders will make decisions that can fully harness a country's potential and maximize its performance.

Secondly, we do not believe that there is a single recipe for reform to be written by social scientists. Any attempt to claim such a one best way and to communicate the "right" reform will always arouse suspicions of ideological motives behind the proposed reforms. Rather, the SGI consist of outcome and procedural indicators that are not linked to distinctive reform philosophies and that are widely accepted beyond political or disciplinary divides.

Our study of executive governance combines state-centric and societal notions of governance, monitors the impact of veto players, and focuses on micro-level functions and processes of governance that reflect a growing consensus on best practices beyond traditional macro-categories of political systems. Our exploration of the survey results in the present paper has produced several important findings:

- The high correlation observable between executive capacity and executive accountability indicates that accountability does not constrain a government's decisionmaking capacity and may even increase executive capacity;
- Countries with particularly high aggregate capacity and accountability scores have established superior arrangements for institutional learning. Countries with weaker executive governance are often characterized by dominant executives and relatively weak non-governmental actors;
- Countries belonging to the Anglo-American tradition of majoritarian, Westminster democracy are characterized by stronger executive capacity whereas the monitoring and information-gathering functions of societal actors are stronger in Central European countries;
- Federal states differ according to whether central governments place more emphasis on ensuring the fiscal and political autonomy of regions or whether compliance with national standards of public services is given priority;
- Top performing countries have developed a combination of strong administrative coordination mechanisms and centers of government with significant policy expertise and steering powers.

These findings constitute only a first analytical effort. Further research will have to examine in more depth whether, how and which disaggregated indicators of executive governance interact with other variables and influence policy performance. The SGI data provide a good basis for such inquiries because, in contradistinction to existing composite indicators and comparative assessments, they explore policy outcomes and governmental practices in greater detail, and they do so for a larger sample of states while using more recent data. In addition, the SGI 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.

Acknowledgements

Several parts of the paper have been intensively discussed in the context of the Sustainable Governance Indicators project. We are particularly grateful to Stefan Empter, Josef Janning, Andrea Kuhn, Leonard Novy and Daniel Schraad-Tischler from the Bertelsmann Foundation who have provided extremely helpful comments and encouragement. We also thank the members of the project's advisory and steering board for their engagement and valuable insights during all stages of the project. Finally, this analysis would not have been possible without the wealth of authoritative background information provided by our country experts. Their detailed reports are highly recommended and can be accessed on www.sgi-network.de.

Bibliography

Afonso, António, Ludger Schuknecht, and Vito Tanzi (2005): Public Sector Efficiency: An International Comparison, in: Public Choice 123, 3-4, 321-347.

Afonso, António, Ludger Schuknecht, and Vito Tanzi (2006): Public Sector Efficiency: Evidence for New EU Member States and Emerging Markets, Working Paper No. 581, January 2006, European Central Bank, www.ecb.org, accessed 1 August 2009.

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

Ben-Gera, Michal (2004): *Co-Ordination at the Centre of Government: The Functions and Organisation of the Government Office*, SIGMA paper 35, OECD, Paris 9 September 2004, accessed.

Benz, Arthur (2003): Konstruktive Vetospieler in Mehrebenensystemen, in: *Die Reformierbarkeit der Demokratie. Innovationen und Blockaden*, edited by R. Mayntz and W. Streeck, Frankfurt/New York: Campus, 205-238.

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

Brusis, Martin (2009): Designing Sustainable Governance Indicators: Criteria and Methodology, in: Sustainable Governance Indicators, edited by Bertelsmann Stiftung, Gütersloh: Bertelsmann Stiftung Publishers, 71-100.

Collier, David, Henry E. Brady, and Jason Seawright (2004): Sources of Leverage in Causal Inference: Toward an Alternative View of Methodology, in: *Rethinking Social Inquiry. Diverse Tools, Shared Standards*, edited by H. E. Brady and D. Collier, Lanham: Rowman & Littlefield, 229-265.

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

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

Esping-Anderson, Gösta (1990): *The Three Worlds of Welfare Capitalism*, Oxford etc.: Polity Press.

Evans, Gord, and Nick Manning (2003): *Helping Governments Keep Their Promises. Making Ministers and Governments More Reliable Through Improved Policy Management*, Internal Discussion Paper IDP-187, World Bank, Washington April 2003, http://siteresources.worldbank.org/INTLICUS/64137337-1094570513360/20250799/Helping%20Governments%20Keep%20their%20Promises.pdf, accessed 1 August 2009.

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

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

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

James, Simon, and Michal Ben-Gera (2004): *A Comparative Analysis of Government Offices in OECD Countries* GOV/PGC/MPM/RD(2004), OECD, Paris August 2004, www.oecd.org, accessed 1 August 2009.

Kaufmann, Daniel, Aart Kraay, and Massimo Mastruzzi (2006): *Governance Matters V: Governance Indicators for 1996-2005*, Working Paper, World Bank, Washington September 2006, www.worldbank.org, accessed 1 August 2009.

Kaufmann, Daniel, Aart Kraay, and Massimo Mastruzzi (2004): Governance Matters III: Governance Indicators for 1996, 1998, 2000, and 2002, in: *The World Bank Economic Review* 18 2, 253-287.

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

Knack, Stephen (2007): Measuring Corruption in Eastern Europe and Central Asia: A Critique of Indicators in Eastern Europe and Central Asia, in: *Journal of Public Policy* 27 3, 255-291.

Kuhlmann, Sabine, Jörg Bogumil, and Hellmut Wollmann, eds. (2004): *Leistungsmessung und -vergleich in Politik und Verwaltung*, Wiesbaden: VS Verlag für Sozialwissenschaften.

Lijphart, Arend (1999): *Patterns of Democracy: Government Forms and Performance in Thirty-Six Countries*, New Haven [Conn.]; London: Yale University Press.

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

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

OECD (2007): *Towards Better Measurement of Government*, Working Papers on Public Governance 2007/1, OECD, Paris, doi:10.1787/301575636734, accessed.

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

OECD/JRC (2008): *Handbook on Constructing Composite Indicators. Methodology and User Guide*, OECD, Paris, http://browse.oecdbookshop.org/oecd/pdfs/browseit/3008251E.PDF, accessed 1 August 2009.

Pierre, Jon, and B. Guy Peters (2005): *Governing Complex Societies. Trajectories and Scenarios*, New York: Palgrave Macmillan.

Pollitt, Christopher, and Geert Bouckaert (2000): *Public Management Reform. A Comparative Analysis*, Oxford etc.: Oxford University Press.

Radaelli, Claudio (2003): The Europeanization of Public Policy, in: *The Politics of Europeanization*, edited by K. Featherstone and C. Radaelli, Oxford: Oxford University Press, 27-56.

Roller, Edeltraud (2006): *The Performance of Democracies. Political Institutions and Public Policy*, Oxford: Oxford University Press.

Schedler, Andreas (1999): Conceptualizing Accountability, in: *The Self-Restraining State: Power and Accountability in New Democracies*, edited by A. Schedler, L. Diamond and M. F. Plattner, Boulder: Lynne Rienner, 13-52.

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

Social and Cultural Planning Office (2004): *Public Sector Performance. An International Comparison of Education, Health Care, Law and Order and Public Administration*, The Hague: Social and Cultural Planning Office.

Tsebelis, George (2002): *Veto Players. How Political Institutions Work*, New York/Princeton: Russell Sage Foundation.

Van de Walle, Steven (2006): The State of the World's Bureaucracies, in: *Journal of Comparative Policy Analysis* 8 4, 437-448.

Wiesenthal, Helmut (2005): *Überlegungen und Messkriterien für den Vergleich von System- und Leistungsprofilen konsolidierter marktwirtschaftlicher Demokratien in der OECD-Welt*, Expertise for the Bertelsmann Foundation, Berlin, mimeo.

Williamson, John, and Stephen Haggard (1994): The Political Conditions for Economic Reform, in: *The Political Economy of Policy Reform*, edited by J. Williamson, Washington D.C.: Institute for International Economics, 527-596.