Van de Walle, S. (2008). Comparing the performance of national public sectors: Conceptual problems. *International Journal of Productivity and Performance Management*, 57(4): 329-338.

COMPARING THE PERFORMANCE OF NATIONAL PUBLIC SECTORS: CONCEPTUAL PROBLEMS

Steven Van de Walle

Associate Professor

Department of Public Administration

Erasmus University Rotterdam

P.O. Box 1738, 3000 DR Rotterdam

The Netherlands

vandewalle@fsw.eur.nl

COMPARING THE PERFORMANCE OF NATIONAL PUBLIC SECTORS: CONCEPTUAL PROBLEMS

Purpose of this paper:

A number of studies and indicators have ranked European countries in terms of the performance of their public sectors. This paper demonstrates there are important conceptual problems with such rankings and comparisons.

Design/methodology/approach:

Using the existing indicators, the paper first shows how European countries' public sectors are ranked. It then goes on to show how conceptual problems with these indicators may lead to incorrect conclusions.

Findings:

Countries' public sector performance cannot be summarised using a single indicator because of our inability to define 'the public sector' and the disagreement on what it means for public sectors to perform.

Practical implications:

Despite increasing demand and supply in policy circles for international public sector indicators, the existing ones are unreliable.

What is original / value of paper?

This paper extends the assessment of international public sector indicators beyond a mere technical evaluation.

Keywords: governance indicators, public sector performance, comparative public administration

COMPARING THE PERFORMANCE OF NATIONAL PUBLIC SECTORS: CONCEPTUAL PROBLEMS

Introduction

The demand for indicators that allow comparing the performance of countries' public sectors internationally is increasing. This trend existed for some time already in the development community, where such indicators are used -or at least referred to- in decisions whether or not to give development aid to certain countries (Arndt & Oman, 2006). The trend is somewhat newer in a European context. In recent years, both OECD and the European Institutions have stepped up efforts to contribute to the debate, and the former's 'Management in Government' project intends to improve the availability of current data.

While some international country competitiveness rankings have included information on the functioning of countries' public sectors for several years, they have always focused on just a very limited aspect of the public sector. Likewise, we have seen a steady increase in the efforts to measure and compare national health and education sectors. Attention for measuring the performance of public sectors in general is much more recent. In Europe, we can trace its origins back to a working paper by the European Central Bank on 'Public Sector Efficiency' in 2003 (Afonso, Schuknecht, & Tanzi, 2003), and subsequent efforts by the Dutch Social and Cultural Planning Office (Social and Cultural Planning Office, 2004). At the same time, several existing data sources, notably two international competitiveness rankings, and the World Bank Governance Indicators (Kaufmann, Kraay, & Mastruzzi, 2006), allow to some extent to rank European countries based on the performance of their public sectors.

The need for such indicators is chiefly motivated by two factors. One is to use the information for naming and shaming countries with a low score and to stimulate them to improve the performance of their public sectors. By making information available, countries would become more aware of their comparative strengths and weaknesses and be challenged to reform. The other factor is related to countries' reputations as efficient or as bureaucratic. Countries may not always deserve their reputation, but these reputations do influence people: businessmen are likely to refrain from investing in countries notorious for their bureaucracy, highly-skilled expats may be unwilling to work in countries with bad health and education systems, and donors are reluctant to give aid to developing countries known for waste and corruption. Public sector performance indicators may help to reassess these reputations.

Problems of data quality with the currently existing indicators have in the meantime become well-documented (Boyle, 2006; Van de Walle, 2006; Arndt & Oman, 2006). They include biased or unreliable samples, an excessive reliance on subjective evaluations, and uncomfortable methods of aggregation. Yet, I argue in this article that problems of data quality are not the most pressing ones, because, with some effort, they can be overcome.

The main argument in this article is that, if we want to compare the performance of countries' public sectors, we have to agree on common definitions. In an international comparative context, we are dealing with public sectors that are organised in quite different ways. Simple concepts such as 'public sector', 'administration', or 'government' may mean quite different things in different countries. Furthermore, to decide whether one country's public sector is performing better than that of another, we need to use a common definition of 'performance'. In current sets of indicators, this conceptual discussion is generally ignored, and the problem defined away.

I will first, however, provide a broad overview of where we stand now, and summarise how European countries' public sectors are ranked in terms of their performance, using the main indicator sets. I continue by discussing the two main challenges in measuring and comparing the overall performance of public sectors. The first is defining the public sector: What is government?

Where does government end and society begins, and how can we define government in such a way that international comparison becomes possible? The second challenge is to determine when a certain government output should be seen as a positive improvement. In other words: what is good performance?

Public sector performance: How do European countries rank?

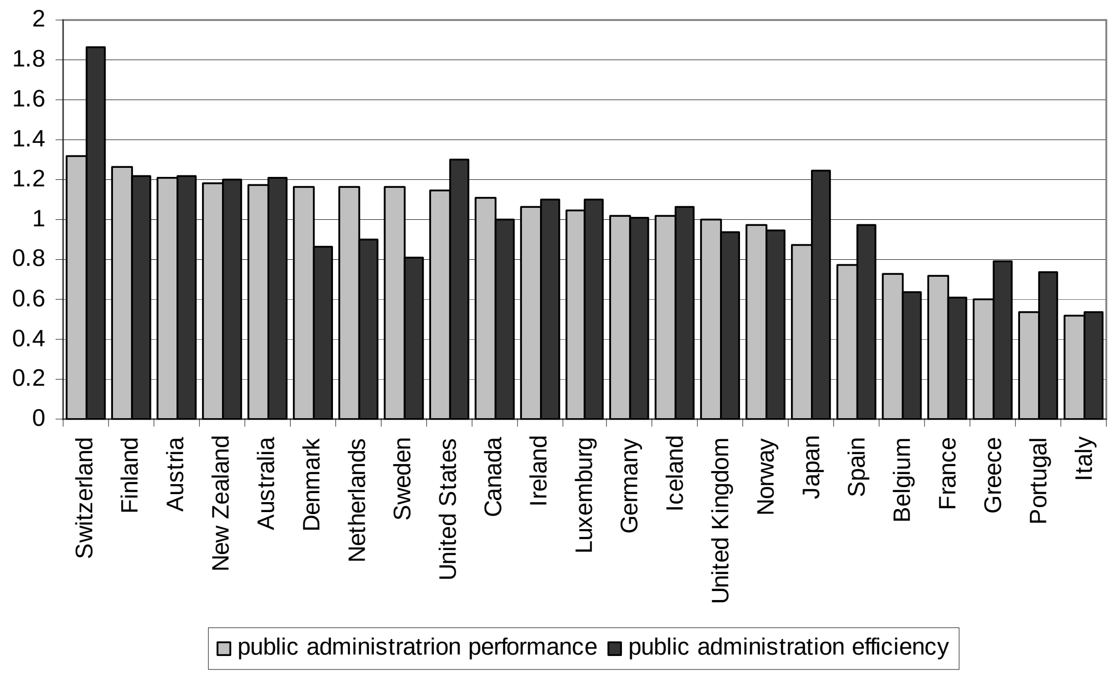
Several studies and indicators have become available in recent years to compare countries' public sectors. Not all of these indicators have been specifically developed to make this comparison, but do contain elements related to public sector performance. In this section, I have a closer look at four of these, and how they rank European countries. Table 1 summarises the characteristics of each of the four indicator datasets. Interested readers are referred elsewhere for a detailed treatment of the strengths and weaknesses of these indicators (Boyle, 2006; Van de Walle, 2006; Arndt & Oman, 2006; Besançon, 2003).

Table 1: Four indicators for comparing countries' public sector performance

Indicator	*Source*	*Country coverage*	*Aim*
Public Sector Efficiency - Administration	European Central Bank Working Paper, 2003: Public sector efficiency: an international comparison	23 (OECD)	Measuring performance and efficiency of the public sector
Government Effectiveness	World Bank Worldwide Governance Research Indicators Dataset	209	Inputs required for government to produce and implement good policies and deliver public goods
Public Institutions Index	World Economic Forum, Global Competitiveness Report (GCR)	117	Determining a country's competitiveness (extent to which government policies are conductive to competitiveness)
Government Efficiency	IMD Business School, World Competitiveness Yearbook (WCY)	61	

A first study is the European Central Bank's working paper 'Public sector efficiency: an international comparison' (Afonso et al., 2003). One of the elements in the comparison, apart from health, education and public infrastructure efficiency, was the performance and efficiency of the public administration. A comparison was made of 23 OECD countries, and resulted in a performance and an efficiency score for these countries. Results are shown in Figure 1. The higher the score, the better the performance and efficiency.

Figure 1: Performance and efficiency of the administration in the OECD countries in 2000

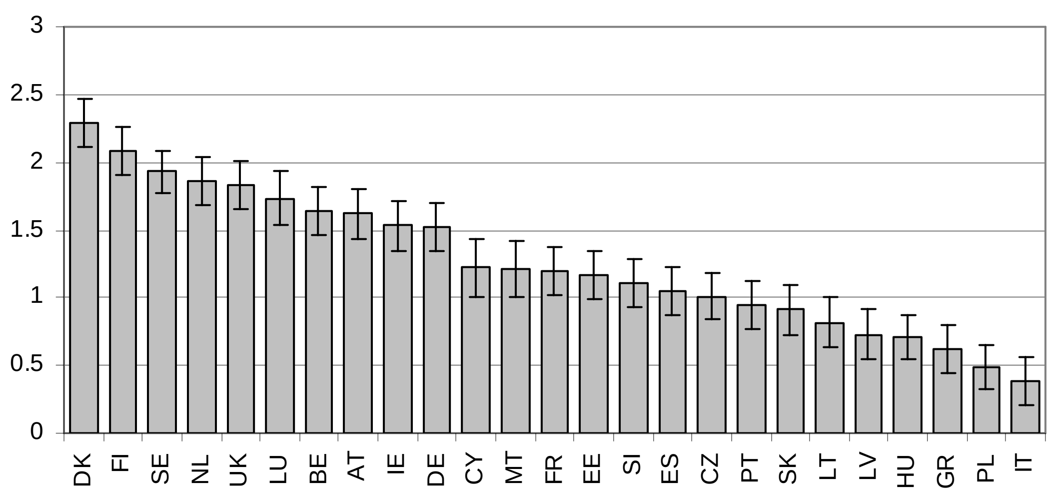


Source: Afonso et al., 2003

Switzerland is the country with the highest public sector performance and efficiency, while Italy is the country with the lowest overall score in the report. This working paper was one of the first attempts at comparing public sector performance and efficiency across countries in a comprehensive way and deserves praise for that. At the same time, however, the study failed to recognise the problems with the data used to measure the performance of public administrations, and it suffered from a small government bias in its efficiency measurement. As these problems have been dealt with elsewhere, I will not repeat them here (Van de Walle, 2006).

A second indicator is the Government Effectiveness indicator, which is part of the World Bank Governance Indicators dataset. This government effectiveness indicator is currently the most comprehensive indicator for assessing the performance of public administrations, even though the indicator is of a quite general and sometimes even vague nature. The indicator currently covers 213 countries or territories worldwide. It measures government effectiveness by looking at 'the quality of public services, the quality of the civil service and the degree of its independence from political pressures, the quality of policy formulation and implementation, and the credibility of the government's commitment to such policies' (Kaufmann et al., 2006: 4). According to this indicator, the Scandinavian countries receive the highest scores, while Italy, Poland, and Greece are dangling at the lower end of the ranking.

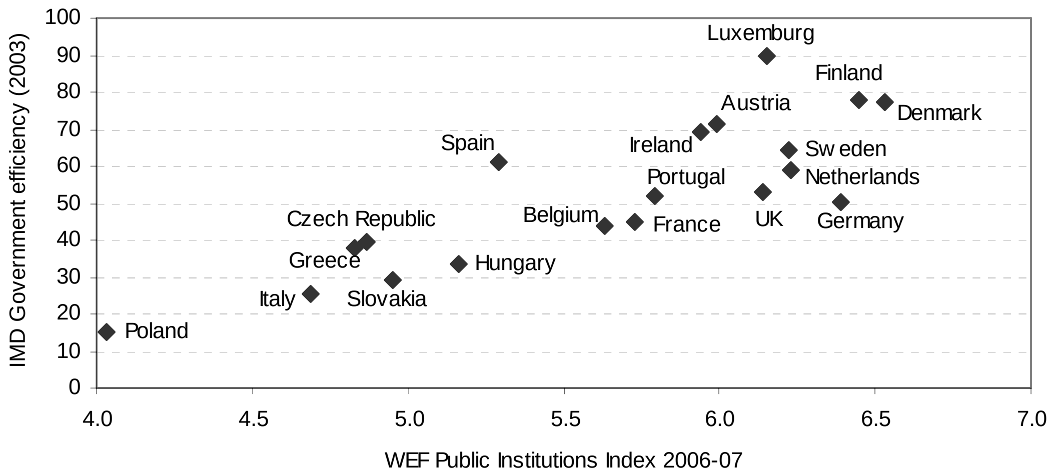
Figure 2: Government effectiveness scores and standard error for EU countries, 2006



Source: World Bank Governance Indicators, 2006. Scores range from +2.5 to -2.5, whereby 2.5 indicates a high effectiveness

A final source of information are the Global Competitiveness Report (GCR) and the World Competitiveness Yearbook (WCY), which have a Public Institutions and a Government Efficiency Index, respectively. These publications are primarily concerned with measuring the competitiveness of a country. A public sector that is conductive to business is an important factor in this competitiveness. The figure shows how the public sectors are scored in 19 EU countries, whereby a higher score means a better performing public sector. The Scandinavian countries and Austria get good scores, while Italy, Poland and Slovakia score quite bad on the public institutions index and the government efficiency ranking respectively.

Figure 3: The performance of the public sector in competitiveness reports



Source: Global Competitiveness Report 2006-2007 & World Competitiveness Yearbook 2004

A problem of definition

The indicators in the previous section appear to offer a straightforward ranking and measurement of countries' public sectors. They do, nonetheless, have a number of technical problems, as mentioned in the introduction. What is however more important is that there also are a number of conceptual problems. Developing a suitable indicator depends on deciding on a number of conceptual and political-philosophical issues. These issues, however, cannot easily be solved. When we want to measure the performance, the efficiency or quality of a public sector, we need to agree on a definition of 'government' and 'the public sector', and we need to find a means to agree on what we consider as 'good' performance.

What is government?

‘The’ quality or ‘the’ performance of ‘the’ bureaucracy is hard to measure, as we are dealing with quite diverse and not clearly delimited concepts. When authors thus refer to the quality or performance of the bureaucracy, they quite frequently use a broad range of concepts. When it comes to measuring performance and comparing performance of different governments, Putnam quite correctly recognised that ‘…because governments do so many different things, they have no single “bottom line”, like profit in a capitalist firm. This fact opens the possibility that different governments might be simply good at different things…’ (Putnam, Leonardi, & Nanetti, 1993: 64). Measuring the performance and effectiveness of government cannot just be achieved by comparing outputs and outcomes to certain objectives, because governments’ objectives are generally vague and necessarily self-contradictory.

‘Government’ is a concept that is particularly hard to define and compare: A Scandinavian-style welfare state is clearly different from a minimal state or an Anglo-Saxon regime, yet we seem to want to directly compare their performance. When we want to compare public sector performance or overall bureaucratic quality, we need to agree on a definition of the public sector, the public administration, and government. Different criteria may be used for determining whether an institution or service can be considered as part of government: financial relations, legal bonds, self-identification, official names and titles, etc. This differentiation is easier in some sectors than in others. How for instance to deal with schools? Or with non-profits that would most likely not exist without the government’s financial support? When public sector efficiency is measured e.g. using the number of civil servants, who do we then count? Only civil servants appointed for life, or also people employed in QUANGOs? There is no, and there can probably be no, agreement on the current size of the government sector (The European Advisory Committee on Statistical

Information in the Economic and Social Spheres, 2003). The popularity of public-private partnerships and agentification has not made this task any easier.

While defining ‘government’ is complicated for researchers, the concept probably becomes even more vague to the participants in the different surveys used for constructing indicators: what should be considered as the public administration or civil service, and what shouldn’t? The meanings of the concepts may differ extensively across countries, thereby endangering comparability. Public administration or civil service may not mean the same to respondents in all countries, and the concepts may sometimes even be unknown or irrelevant.

Determining where ‘public administration’ ends and where ‘government’ begins, or where ‘government’ ends, and ‘society’ begins is not an issue that can be solved by developing better definitions. In some cultures these distinctions may even be considered as entirely irrelevant. Blondel, when speaking about the border between politics and administration even wonders ‘Does government have to have an end?’ (Blondel, 1982: 206). He considers defining government as one of the key problems in comparative research: ‘But we are also hampered in our attempts at general surveys by our gross inability, so far, even to try and give a definition of what government is’ (Blondel, 1982: 5). He identifies ‘where does the government end and the administration begin’ as a basic question (Blondel, 1982: 206). On top of these concerns comes the problem of translation, where *administration, administratie, regering, government, gouvernement, Verwaltung, Regierung, bureaucracy* all have similar but not equal meanings. But this is not just a matter of finding the right translation (Rutgers, 2004). Finally, there is Peters who raises the fundamental question whether there is ‘*a* bureaucracy in most societies, or simply a collection of many bureaucracies?’ (Peters, 1992: 89). Defining the boundaries of government is not just a matter of empirical observation and definition, but also a matter of appropriateness: where do I want government to end?

As a result, we find that the existing indicators do not focus on government or the public sector as a whole, but rather on a specific subsection of government. Many indicators have a quite specific approach to bureaucratic quality, and tend to focus on specific sub domains and -processes of public administrations, rather than on public sectors *per se*. In addition, this selection often contains a normative element. 'Government' is thus conceptually reduced, and this reduction depends on the specific purpose of the researcher when constructing the indicator: In risk indicators, a qualitative bureaucracy is one that is predictable. For donors, it is a bureaucracy where corruption is absent. For comparative political scientists, political-administrative relations are a key dimension of administrative performance. This is directly related to the next question: what does it mean when we say a government is performing?

What is good government? When does government perform?

A second major challenge is related to the definition of 'good' or 'better' performance. Researchers are interested in determining whether government A is performing better than government B. A government that is wasting resources is not performing adequately. One that is not wasting resources is a performing organisation. But how do we determine whether a government is wasting resources? Obviously, we do not only want government to do the things right, we also want it to do the right things. But what are 'the right things'? And how do we determine whether they are doing the right things and that they are done right? This is more than just a technical question. Obviously, there is no market value for measuring government outputs against (Handler, Koebel, Reiss, & Schratzenstaller, 2005: 17).

The apparently easiest approach is to be selective. Rather than measuring all aspects of government performance, we can define a bottom-line: *It doesn't really matter what a government is doing as long as it does X, or doesn't do Y*. This is an approach that is quite popular in research. The need to have an simple indicator that is easy to handle in research has lead to the use of very specific indicators. Corruption is a good example. Corruption is used as thé indicator *par excellence*. A country's institutions receive good grades when they are found not to allow corruption. This obviously is a quite narrow interpretation, and many will agree that the mere absence of corruption is by no means a guarantee for high performance. Another quite popular approach is to reduce government performance to being conductive to business: good governments are governments that support business by limited regulatory intervention, low taxes, and a high degree of stability or absence of political and administrative risk. Red tape, in this view, is inevitably related to inefficiency. Three of the indicators in the first section of this article fit within this tradition.

What the indicators generally fail to acknowledge is the fact that governing implies choosing between incommensurable aims and determining which side of a dilemma should receive more attention. A 'good bureaucracy' for business, may be a bad bureaucracy for the development sector and vice versa. Many indicators promote a certain vision on the role of government in society. This should not come as a surprise, because the indicators were developed to be used in quite specific contexts. Several indicators grew out of a number of risk indicators which were developed to guide companies in investment decisions. It is in this context perfectly legitimate to measure the degree to which the regulatory environment represents an obstacle to business. The risk rating agencies tend to adhere to a non-interventionist ideology. La Porta et al. explicitly define good public sector institutions as institutions that 'are instrumental to economic growth' (La Porta, Lopez-de-Silanes, Shleifer, & Vishny, 1999: 222). They do, however, recognise that extensive government intervention does not equal inefficient government. Similar biases also exists for indicators used by international donors, indicators which are often evolving and normative (Besançon, 2003: 3).

Government is often not the only actor responsible for realising a certain effect. Even though outcomes can sometimes be measured, it is not always easy to distinguish whether an indicator's evolution should be considered a 'positive' one. Public organisations have ambiguous goals (Fukuyama, 2004), and reaching one goal may happen to the detriment of another. Definitions of effectiveness in a sector may be disputed. When is a government successful? Should it aim to increase Human Development (UN), or should it primarily focus on fighting extreme poverty (World Bank)? Or perhaps reaching the Lisbon criteria (EU) is government's primary responsibility? However desirable, certain public investments may be considered unacceptable for reaching certain targets in some countries, while they may be perfectly acceptable in other. Additionally, and related to the previous section on defining government, it is hard if not impossible to determine how much of an outcome is due to government policy.

Finally, we may wonder whether subjective assessments which are at the basis of most indicators give us an adequate picture of the performance of the public sector. All indicators that were used in the first part of this article rely to a large extent on subjective data. These mostly are surveys of citizens, executives or experts. Subjective perceptions, however, do not necessarily reflect actual performance, and may thus return an inaccurate image of public sector performance (Goodsell, 1983). This need not be problematic. Very often, there is a high correlation between the different country ranks using different indicators. Yet, this need not necessarily be seen as a sign of their reliability. Rather, expert opinion on a broad range of public sector related issues may 'simply reflect some general underlying sentiment toward a country' (La Porta et al., 1999: 234). Subjective indicators may only measure the public administration's image, rather than being an evaluation of performance. Opinions may reflect historical experience rather than current performance, or may reflect exceptional encounters rather than average service. The indicators I have discussed at the beginning of this article all depend to some extent on experts' opinions.

Experts' opinions, however, may be affected by the ideology of those who have devised the indicator (Knack & Manning, 2000: 13; Malik, 2002: 5).

Conclusion

Demand for indicators that allow for comparing the performance of public administrations internationally is increasing. A diverse range of actors is interested in indicators for measuring the quality of bureaucracies at the macro-level, rather than at the micro-level of specific services. These actors include inter- and supranational organisations, the international donor community, and scholars. Demand in comparative public management remains rather limited, but it is to be expected that public management scholars will speed up their efforts to compare the performance of countries' administrations.

The supply of indicators is also rapidly increasing. A growing number of datasets is being developed and promoted, with the World Bank governance indicators as one of the best-known examples. Yet, many of these indicators and datasets suffer from conceptual and data quality problems, making them less suitable for comparative public management scholars. I have argued in this article that, despite the fact that there are still many quality problems, the main problem with developing indicators for measuring and comparing the performance of public sectors at large is not a technical problem, but a conceptual one. Measuring the performance of government or the public sector requires we have a definition of government or the public sector, and that we can determine where government ends and society begins. With many new and old types of private and citizen involvement in government, and with different roles and positions of government in society, such a definition is not straightforward. Even when we would be able to define 'government', there would remain discussion about when its overall performance was to be considered good or improving. On top of this come difficulties in attributing certain social outcomes to the intervention of government.

There is a real risk that the indicator boom may lead to superficial international comparisons. When comparative public management wants to build a research programme around the comparison of the performance of public administrations, it can build on the work that has been going on for decades in other disciplines. Blind reliance on existing data, however, is likely to endanger the quality of research. Public management scholars need to adapt the existing indicators to new developments in the definition of government, and need to raise the methodological standards of some of the older datasets.

And the same goes for policy makers. Every time a new indicator or ranking is released, the results are taken at face-value: our country's public sector is doing better/worse than that in country X. I hope to have demonstrated that such comparisons are good for putting the issue of service improvement on the public and political agenda, yet that they tell us very little about the performance of the public sector.

Reference List

Afonso, A., Schuknecht, L., & Tanzi, V. (2003). Public sector efficiency: an international comparison. European Central Bank Working Paper no. 242.

Arndt, C., & Oman, C. (2006). Uses and abuses of governance indicators. Paris: OECD Development Studies Centre.

Besançon, M. (2003). Good governance rankings: The art of measurement. WPF reports no 36 .

Blondel, J. (1982). The organization of governments: A comparative analysis of governmental structures. London: Sage.

Boyle, R. (2006). Measuring public sector productivity: Lessons from international experience, CPMR Discussion Paper no.35. Dublin: Institute of Public Administration.

Fukuyama, F. (2004). Why there is no science of public administration. Journal of International Affairs, 58(1), 189-201.

Goodsell, C. T. (1983). The case for bureaucracy: a public administration polemic. Chatham: Chatham House Publishers Inc.

Handler, H., Koebel, B., Reiss, P., & Schratzenstaller, M. (2005). The size and performance of public sector activities in Europe. WIFO Working paper 246/2005.

Kaufmann, D., Kraay, A., & Mastruzzi, M. (2006). Governance Matters V: Governance Indicators for 1996–2005. Washington, D.C.: World Bank.

Knack, S., & Manning, N. (2000). Towards consensus on governance indicators: selecting public management and broader governance indicators.

La Porta, R., Lopez-de-Silanes, F., Shleifer, A. , & Vishny, R. (1999). The quality of government. The Journal of Law, Economics and Organization, 15(1), 222-279.

Malik, A. (2002). State of the art in governance indicators. Background paper for the Human Development Report 2002, Occasional paper 2002/7.

Peters, B. G. (1992). Comparative perspectives on bureaucracy in the policy process. L. B. Hill (ed.), The state of public bureaucracy (pp. 87-110). Armonk, New York : M.E. Sharpe.

Putnam, R. D., Leonardi, R., & Nanetti, R. Y. (1993). Making democracy work: civic traditions in modern Italy. Princeton, New Jersey: Princeton University Press.

Rutgers, M. R. (2004). Comparative public administration: navigating Scylla and Charybdis - global comparison as translation. Administrative Theory and Praxis, 26(2), 150-168.

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.

The European Advisory Committee on Statistical Information in the Economic and Social Spheres. (2003). The size of the government sector: how to measure. 24th CEIES seminar, Vienna, 23-24/10.

Van de Walle, S. (2006). The state of the world's bureaucracies. Journal of Comparative Policy Analysis , 8(4), 439-450.