

Van de Walle, S. (2009). 'International comparisons of public sector performance: how to move ahead?' *Public Management Review*. 11(1): 39-56

INTERNATIONAL COMPARISONS OF PUBLIC SECTOR PERFORMANCE: HOW TO MOVE AHEAD?

Steven Van de Walle*

Associate Professor

Department of Public Administration

Erasmus University Rotterdam

The author would like to thank the reviewers and the participants in the panel on 'The Metrics of Tracking Change in the Public Sector' at IRSPM XI for their helpful comments.

INTERNATIONAL COMPARISONS OF PUBLIC SECTOR PERFORMANCE: HOW TO
MOVE AHEAD?

Abstract

Measuring and comparing the overall performance of countries' public sectors requires agreement on definitions and objectives of government. I argue that such an agreement is about finding a consensus rather about finding *better* definitions. Measuring government requires a number of *leaps of faith*, where certain definitions, assumptions and statistics are accepted as *good enough* for measurement and comparison. The political science and economic research community have a different tradition of dealing with such agreements and leaps of faith, and this is reflected in their approaches to measuring and comparing the performance of public sectors. The implications of these traditions are particularly visible in the usefulness of measurement and indicators for policy makers.

Keywords: Performance measurement, size of government, philosophy of science, public sector statistics

Introduction: An indicator problem which really is a conceptual problem

At an ESRC conference *Do you get what you pay for? Getting to grips with public service productivity*, in London in September 2006, the conference chair, Christopher Hood, asked the presenters whether they saw the main problem with current measurement of public sector productivity as a conceptual or as a measurement problem. The economists claimed it was a measurement problem, while the public administration scholars and political scientists claimed it was a conceptual problem. And this very well summarises where we currently stand in the debate and scholarship on measuring and comparing the performance of countries' public sectors.

Certain European countries are known for their hopelessly bureaucratic public sectors, while the public sector in other countries is generally seen as innovative and efficient. Different sets of competitiveness indicators and governance indicators generally confirm these popular conceptions and national stereotypes (Van de Walle, 2006). Yet, solid international comparative empirical evidence is generally not available (O'Mahony and Stevens, 2006), and the indicators that do exist are often not reliable or conceptually flawed (Van de Walle, 2006). Aggregate measurement of public sector performance is receiving increasing attention in certain sectors such as health and education (Kuhry, 2004; O'Mahony and Stevens, 2006), but attention for the administrative-bureaucratic part of government, or even for government in general has remained rather limited.

The efforts to improve measurement of *government*, e.g. in a Eurostat or OECD context, have tended to focus on technical aspects of measurement, and often ignore the conceptual discussion about *government*. They take *government* and its objectives as a given. The argument I will develop in this article is that, in the international measurement of *government* we are faced with a conceptual problem, not with a measurement problem, and that to measure and internationally compare government and government performance, there needs to be an agreement on definitions of government and on governments' objectives. Both things cannot be done without excessive simplification.

I first deal with the question *what is government*, and then proceed to the question *what should government* do. In a further section, I focus on the problem of attribution and causality in

measuring government performance, and in the final sections I discuss implications of these questions for measuring government and comparing government performance internationally.

What is government?

One of the key conditions for performance measurement is that you know what you are measuring (Radin, 2006: 207-8). To measure and internationally compare the performance of government, you need to be able to define *government*. This simple word *government* is actually a very vague concept that is interpreted in different ways, especially when we move across cultural boundaries. Americans and Europeans, for instance, have clearly different conceptions of *the state* (Rutgers, 2001). *Public administration* is also seen in different ways, and where many Europeans would regard it as an institution, Anglo-Saxons would in the first place think about the discipline or the activity.

The role of third sector organisations, quangos and private partners in government, a role which is by no means new, does not make defining government any easier. A mix of traditional ministries, part publicly-owned private companies, third sector organisations, regulated industries etc., does not make distinguishing between the public and private very easy, if not impossible (Pollitt, 2003: 12-14).

Determining the size of government is not straightforward (The European Advisory Committee on Statistical Information in the Economic and Social Spheres, 2003; Handler, Koebel, Reiss and Schratzenstaller, 2005). Indicators such as public sector employment, public sector expenditure, and tax income are less clear-cut than they appear to be. I give just one example related to public sector employment. Determining the size of public employment requires a definition of what public employment is, and this definition needs to be relevant and comparable for all countries included in an international comparison. Counting civil servants employed in a traditional civil servant statute is easy, but how to deal with those working at the edge of government? Does medical staff employed in private or non-profit hospitals have to be included when their wage is in the end paid through a public health insurance system? Differences in the organisation of health systems have an important impact on the recorded size of public employment. Making doctors

private consultants instead of government employees hardly changes their job, but it does create a challenge for public employment statistics. How are such statistics influenced by decisions to replace labour-intensive treatments by capital-intensive treatments, and vice versa? (Recall Baumol's comments on productivity growth in labour productive and capital intensive sectors (1967)) Do we include cleaners at government offices, and do we still include them when the service becomes outsourced? Do we include parish priests when their wages are paid by government? Do we include staff of an environmental NGO when most of this NGO's budget comes from government subsidies? And what about aerospace industry which largely depends on government funding (Bozeman, 2004: 7-13)? And how do such statistics deal with multi-level governance arrangements? If certain government tasks are performed at different levels of government, how do we then add up the numbers? How can we compare a country where certain government tasks are performed by the national civil service to a country where such tasks have been transferred to local or to supra- or international bodies?

Government and public administration are part of and interact with society. There is no clear boundary between public administration, politics, and society. This has implications for definition and measurement (Blondel, 1982). In attempts to measure government, we prefer to work with well-defined objects of measurement, which the public sector clearly isn't. The public sector is not an integrated and clearly delineated entity. The way of looking at the interactions between government and society, and accepted interpretations of where government ends and society begins may be quite different across countries. As a result, in international comparisons of public sector performance we can only pretend to know what we are measuring.

What should government do?

What are government's objectives?

In a private environment, money value can be used to determine the value of output, and for private companies, 'the value of product manufactured is a valuationally neutral element' (Simon,

1976: 175) -- but even this can be disputed. Not so for public services. For public services a substitute for money value of output must be found, which, according to Simon, can be acquired by 'a statement of objectives of the activity' (Simon, 1976: 175). He admits this is not simple task. Objectives should be stated in final ends, but these values are 'seldom expressible in concrete terms' (Simon, 1976: 176). The problem with this is that an activity may realize two or more values, making the attribution of weights difficult (Simon, 1976: 176). He states that this attribution of weights is done *somewhere, sometime in the administrative process*, but this does not always happen consciously and deliberately, but is implied in the decisions taken. Additionally, the degree to which an objective is to be reached needs to be determined.

While Simon recognises some problems in this approach, he remains rather positive about the possibility of determining government objectives and measuring the extent to which these objectives are reached. What he proposes is actually the recipe proposed by most organisational consultants: determine the mission of your organisation, set objectives and make sure that all of your organisation's activities contribute to these objectives. But it is exactly this objective setting that is the hard part. How can we decide on objectives in a political system?

Economists would suggest looking at citizen preferences and social production functions but this requires we accept an assumption that value preferences remain more or less fixed, lest we not be forced to constantly change government's objectives, making effective measurement impossible. This approach places the citizen-as-consumer at the centre of evaluations of government, and has difficulties dealing with radically different views in society of the same problem. This requires governments to accommodate these different views by ignoring them, creating a consensus, or by influencing preferences. The strategy followed depends on changing power balances in society. In a public sector context, which is by definition a political context, it would also be unwise to take the objectives of a policy or of government as stated by all or some policy makers as the true objectives. Objectives may not be revealed, objectives may be contradictory, or policy makers may just not know what exactly their objectives are.

‘The measurement of efficiency cannot be more valid than the process by which objectives are defined’ (Simon, Smithburg, and Thompson, 1959: 506), and by saying this Simon correctly identifies the main problem in measuring and comparing government performance. Value conflict is an essential feature of governing. Governing is about dilemmas and choice, and measuring government performance requires that these dilemmas are first solved. Performance measurement is essentially based on a teleological worldview. Measurement is only possible in a situation where we accept to work with a number *ceteris paribus* assumptions and a number of leaps of faith. Measuring government-wide performance requires that we agree on objectives to measure performance against, and that we accept certain imperfect data as good enough for measurement (e.g. GDP, public employment statistics, micro-performance indicators). Econometric modelling of government performance requires we act as if the basic problem *what is government/what should government do* is already solved. This is a very pragmatic approach.

The quest for a consensus on a single yardstick

When we want to know whether governments perform, we need to decide on performance criteria. Different yardsticks for overall government performance have been suggested, but none of these is at the same time undisputed and specific enough to be used as an outcome criterion. Using a single criterion to evaluate governments is of course a denial of governments’ breadth and variety, and one could say that ‘the idea of deriving a single measure of productivity for the nation, a sector or an organisation is unrealistic’ (Boyle, 2006: 34). In some sectors of government, outcome indicators with broad yet not universal acceptance have been developed, such as life expectancy in health, or PISA scores in education. In other areas there is considerably more discussion.

Definition of desired outcomes is essentially a political act. *Good governance* has become an especially popular concept, and is seen as a widespread consensus of what constitutes best practice in the functioning of public administrations (Brusis and Ramel, 2006: 3). The prevalence of indicators suggests there is a consensus about what constitutes good bureaucratic governance.

This *consensus*, however, is far from a consensus, and merely reflects dominant values, policy fashions, or mere convenience. In most definitions, good governance refers to ‘a tripartite set of competing conceptions of what constitutes “good” government’, containing ‘the rule of law, democracy, and efficiency/effectiveness’ (Rothstein and Teorell, 2005: 17). Despite attempts to present *good governance* as a neutral concept embodying ways of working and outcomes on which everyone would agree, it is essentially a political concept. In the international donor community, using *good governance* allowed to frame sensitive questions ‘under a relatively inoffensive heading and usually couched in technical terms’ (Hewitt de Alcántara, 1998: 107). This crucial vagueness of the concept of *good governance* (Doornbos, 2003; Boas, 2001) helped to put the concept on the agenda (Bovaird and Loeffler, 2003), but does hinder adequate measurement and rigorous analysis. *Good governance* has many dimensions and is sometimes highly prescriptive (Grindle, 2004; Bovaird, 2005). While presented as a neutral concept, definitions of *performance* in the public sector often reflect a certain paradigm (Santiso, 2001; Grindle, 2004; Cirillo, 2006). In a government context, and in evaluating the effectiveness of government organisations or government actions, we need agreement on what *counts as a result* (Pollitt, 2000: 187).

Using weights

A second remark concerns the use of weights. To establish overall performance or productivity, we need a way to attribute weights. Giving all factors equal weights does not solve the problem, as this is just a special case of attributing weights (Cherchye, Moesen, Rogge, and Van Puyenbroeck, 2007: 114). Governments have multiple outputs, and these all in their own way contribute to the output of *government* as a whole. Government output at the macro level cannot be considered a simple aggregation of government outputs at the micro-level. In other words, achieving one outcome interacts with achieving other outcomes, and ‘[I]t is asserted that productivity cannot be measured in a meaningful way when there is more than one output’ (Murray, 1987: 171). This attribution of weights again requires that we agree on governments’

objectives, and the relative importance of these different objectives. It is however unlikely that we can establish unique weights for a society. Rather, it is more likely that different groups or stakeholders would prefer different sets of weights. Deciding on a single set of weights is ultimately an arbitrary exercise, especially in international comparisons. Believing that a stable set of unique weights could be agreed upon would make the government or society that is being measured a-political.

Can government ever reach its objectives?

Performance improvement and the problem solving metaphor

In the indicator culture, it is expected from policy makers that they can present their strategy and desired outcomes in simple terms, and that they can link these outcomes to indicators. Being able to do so gives managers an aura of competence and mission. Governing, in such an approach, is presented as an unproblematic act. Yet, policy making is an untidy process (Lindblom, 1980), and not a rational, ordered one. Governing is more than *solving problems*, yet there seems to be an entrenched belief that there are clearly definable courses of action that will lead to *better* outcomes (Lindblom and Cohen, 1979). Boyle, talking about governments' obsession with measuring, describes this traditional view of measuring as one where counting is seen as a way to improve the world (Boyle, 2001: xiii). As we will see later, this thinking reflects the influence of the natural science model in the study of government. Flyvbjerg refers to the social scientists in the 1960s and '70s who claimed that 'if natural science and engineering could put a man on the moon', social scientists should also be able to solve a number of social problems (Flyvbjerg, 2005: 39).

Talking about *better* outcomes has as an *advantage* that it drives out contradictions and 'pursues efficient solutions to messy social problems' (Adams and Balfour, 1998: 139). This vision of government reflects 'the continuing influence of the "modernist" faith in progress informed by reason' (Sanderson, 2002: 1). Oakeshott described the Rationalist, as someone who cannot imagine 'politics which do not consist in solving problems, or a political problem of which there

is no “rational” solution at all’ (Oakeshott, 1991: 10). In most societies, political consensus is more likely to be absent than present, and very often we don’t agree what problems are and what would be a good solution (Adams and Balfour, 1998: 137). Or, in Popperian terms: There is no generally accepted problem-situation (Popper, 2002: preface to the first edition 1934).

Incompatibilities

The attractiveness of presenting government as a problem solver with clearly defined problems to be solved, or objectives, is that it allows measuring the achievement of these objectives. What is often forgotten however is that government is constantly dangling in an uneasy equilibrium between competing values (Adams and Balfour, 1998: 142; Kaufman, 1956; Rosenbloom, 1983). As a result, Olsen argues, there is *‘no best way of organizing public administration so that it is always most helpful for citizens’* (Olsen, 2004: 69). Administrative reform and improvement is therefore always relative: ‘Improvements in one area may be offset by deterioration elsewhere. Even if administrative performance is improved, other values -- equality, democracy, security -- may be sacrificed’ (Caiden, 1968: 348). At the level of the internal functioning of public organisations, Simon (1946) showed that ‘the traditional principles of public administration could be arrayed into logically contradictory sets’ (Ostrom, 1974: 7). The pursuit of defining government objectives and measuring government’s progress towards these objectives reflects a firm belief in a Platonic ideal of truth where there is one true answer, and where the rest are errors; that there is a dependable path towards the discovery of these truths; and, finally, that these truths must ‘be compatible with one another, and form a single whole’ (Berlin, 1991: 6).

Incompatibility of different preferred objectives is a common feature in any society and political system. This makes the development of indicators very difficult. If certain values may not be compatible, the attainment of which ones will we then measure? Even if we would decide to choose, we need to realise that given multiple objectives, all performance measures will be interrelated (Martin and Smith, 2006). Organisations cannot ‘attend to all of their goals simultaneously’ (March, 1988: 3; Cyert and March, 1963), and this *failure* will have an impact on how goal achievement will be reflected in the indicators.

Ambiguity as strategy

Organisations, and thus also public sector organisations, are characterised by ambiguity. They have ‘inconsistent and ill-defined objectives’ (March and Olsen, 1976: 12). Public bureaucracies are often asked to pursue very ambiguous tasks (Wilson, 1989). Having clear and undisputed objectives as a government may be seen as a desired path, and befitting a rational organisation model, but ambiguous or non-manifest goals may actually be preferred. Political organisations are built on the principle of conflict, and ambiguity can actually be very helpful in governing because it helps to reduce tensions. Remaining vague about strategy may be beneficial (Quinn, 1980; Eisenberg, 1984). Symbolic action is sometimes all it takes to reduce tensions.

But at the same time, organisations are also bodies where criticism of the status quo is cultivated (Brunsson, 1989: 24). This constant scepticism makes it difficult to determine objectives, as they are likely to change all the time. Measurement is difficult in a situation where there is constant conflict and competition between values. In addition, organisations do not only produce actions that contribute to objectives, and ‘one of the ideological outputs of organizations is talk’ (Brunsson, 1989: 26). How does measurement deal with talk? Symbols, values and discourse are important government outputs, and ignoring them in the measurement does not make them disappear. But it would equally be wrong to take the discourse as a perfect indication of governments’ objectives. Measuring government outputs and outcomes against the talk government creates is very likely to lead to a very low performance score.

Effectiveness and problems of attribution

Attribution and causality as a belief

When we want to evaluate government’s actions we not only need to know what its objectives were in the first place, but also how government itself has contributed to achieving those objectives. Even when we would be able to decide on government’s objectives, there is a problem of attribution: ‘In distinction to outputs, outcomes cannot usually be simply attributed to

government actions or processes' (Bertok, Hall, Kraan, Malinska, Manning, and Matthews, 2006: 9). This problem is already very visible in the case of minor policies: was the drunk-driving campaign the reason for the decline in road accidents? The problem becomes huge when moving to a more aggregate level, as is the case in this article on measuring and comparing the overall performance of countries' public sectors. Changes in the aggregate performance indicators currently in use tell us very little about the government actions and activities supposedly lying behind them (Van de Walle, 2006).

The problem of attribution becomes a fundamental problem where public sectors are moving towards a greater use of broad outcome indicators in making and evaluating policy, such as well-being, sustainability, life satisfaction, 'community' or even happiness (Bertok et al, 2006). Some, for this reason, advise against using outcome indicators: 'Putnam (1993) rejects the idea of including outcomes in productivity measurement. His argument is that to focus on outcomes [...] includes changes over which the government has no control' (Boyle, 2006: 5).

We have a strong desire to attribute certain changes in the indicators to certain government actions. That changes may have occurred *despite* government intervention, or even just by chance (Bonar Blalock, 1999: 118), is often hard to swallow. We often need to believe that certain things have made a difference. Pollitt for instance demonstrated this by evaluating the effects of NPM reforms, where there has been a tendency to overstate changes, and to attribute many changes in the public sector to the implementation of an NPM-style programme (Pollitt, 2000). Attributing government's actions to certain outcomes requires a leap of faith, where a conscious or unconscious decision is taken to attribute certain outcomes to certain government actions. There needs to be a belief of causality. Having a causal story accepted by others is an important political instrument (Stone, 2002: 203). It allows us to assign responsibility for certain problems, whether or not there is good evidence of a causal relation (Stone, 2002: 189).

The problem of service chains

Often, 'performance is achieved in a network of dependencies' (De Bruijn, 2002: 36). Most public services are organised and delivered by a multitude of actors, many but not all of which are public. The delivery of services such as health, transport, or education is based on a service chain, where many interlocking actors interfere. A patient visits a GP, is sent to a specialist, hospitalised, sent home and taken care of by social services or by family. Where and how do we then measure government performance in delivering this health-related service? We could measure the performance of the hospital, but by doing so we ignore those elements of the health service chain that precede and follow the hospitalisation. The organisational reality does not necessarily correspond to the service reality, and it is hard to determine where the chain begins and ends. How does the measurement account for the difference between a patient visiting a GP and getting referred to hospital, and a patient that directly goes to hospital? Does measurement of performance end when the patient leaves the hospital and recovers at home? Another example would be where we want to measure government's performance in combating crime: how far up or down the chain do we want to go in this measurement? Do we include the police and prisons? Obviously we would. But what to do with social workers' prevention work?

Given that most public services are delivered in interaction between government and society, and in interaction between provider and client, defining where a service *ends* is difficult. Where a service *ends* is generally different from where measurement ends. Deciding on the *end* of a service is a measurer- or policy maker-imposed delineation, not an experiencer-imposed one.

When service providers use public sector performance measurement as policy or control mechanism, this vagueness of how to define a public service can easily be exploited by shifting costs to those components of the service chain that fall outside the policy-maker imposed delineation (Bevan and Hood, 2006). Patients can be sent home early to recover, allowing hospitals to shift costs to families or social services.

How then to move ahead?

Measuring overall government performance is either built on beliefs and arbitrary decisions on what government is, on what government's objectives are, and on how causality works, or it can't be done. Assumptions, beliefs, and leaps of faith are essential wherever measurement is done. The extent to which we are willing to accept certain assumptions and necessarily arbitrary decisions to advance the *measurement* of government depends on our conception of social science.

The function of composite indicators

Useful measurement does not always mean correct measurement. In the public sector, we want indicators that correctly reflect that what is being measured, yet that are at the same time sufficiently simple and attractive to be useful. Performance indicators need to reflect the diversity and complexity of what is measured, yet at the same time they have to be kept very simple for accountability purposes (De Bruijn, 2002: 5). It is difficult, if not impossible to develop an indicator that is perfect for all uses and objectives (Behn, 2003). Many composite indicators are developed for their advocacy value:

One may note that the controversy on the use of statistical indices unfolds along an analytic versus pragmatic axis. There is abundant literature on the analytic problems associated to even well-established statistical indices such as GDP [...]. This literature hardly seems to dent GDP's rather universal pragmatic practical acceptance (Saltelli, 2007: 68).

In other words, an indicator does not have to be perfect to be useful or to be used. As Bertok et al. remarked about measuring government outcomes, 'outcome indicators have a role not just in focusing political debate, but in broadening the involvement of the wider public in policy-making' (Bertok et al, 2006: 20). When indicators are used as a rhetorical tool for agenda setting, these same indicators should not be used to measure performance because they were not in the first

place invented for this reason. Not reaching the ambitious targets is then indeed a very likely outcome, and measurement can only lead to *bad* results.

Before we can develop better indicators to measure and compare public sectors internationally, we need to decide why exactly we want these composite indicators. Given the OECD's lead in current efforts, improvement of the public sector's functioning rather than academic perfection is the likely objective of measurement. Our approach in this article however is on how we can develop better indicators to reflect the complex reality of government performance. In the next section we compare two possible paths for progress.

Leaps of faith? Economists and political scientists leap in different ways

In this article, I have thus far demonstrated that government, government's objectives, and the causalities between government actions and societal outcomes cannot be defined in an undisputed way. For measurement to be possible, working definitions and assumptions need to be established. This procedure requires a number of leaps of faith, ignoring known complexities. This is inherent in all scientific progress, as demonstrated by Popper's take on the role of *faith*:

[...] I am inclined to think that scientific discovery is impossible without faith in ideas which are of a purely speculative kind, and sometimes even quite hazy; a faith which is completely unwarranted from the point of view of science, and which, to that extent, is "metaphysical" (Popper, 2002: 16).

In the current scholarship focusing on comparing and measuring government performance, we see two different approaches to the desirability of taking such a leap of faith. These two approaches reflect two different paths of progress in studying overall government performance. One is that propagated by econometricians, who have focused on measuring *government* in their models for e.g. explaining economic growth. More specifically, they have focused on measuring the quality

of (public) institutions. While they often acknowledge problems with the existing indicators (Olson, Sarna, and Swamy, 2000), this is not always the case. This has however not stopped them from developing and testing new models. Political scientists, on the other hand, have been much more hesitant using such indicators and comparing governments at the aggregate level. Instead, they have tended to rely on either country case studies or detailed comparisons of elements of governments rather than governments in general. In the discipline very little agreement exists on the definition and measurement of even the most fundamental concepts. As a result, we have not seen considerable progress towards general government measurement, or towards the international comparison of government performance. It is no surprise that most if not all of the recent efforts to measure and compare overall government performance has been initiated by economists and econometricians (Kuhry, 2004; Afonso, Schuknecht and Tanzi, 2003; Afonso, Schuknecht, and Tanzi, 2005).

These different approaches present us with two different paths for progress in measuring and comparing overall government performance. In a way, they appear to reflect the difference in how Kuhn and Popper described how science works or ought to work. The basic element of this discussion is on whether working within an established paradigm contributes to scientific *progress*. We summarise their differences in the table.

Thomas S. Kuhn	Karl Popper
The Structure of Scientific Revolutions (1962)	<i>Logik der Forschung</i> (1934), translated as The logic of scientific discovery (1959)
Science works within a paradigm, it agrees to common patterns of work	Science as a process of constant challenges, <i>falsification</i> , not working within paradigms
subjectivist, relativist, historicist; scientific pluralism	objectivist, realist, positivist; unified conception of science

This difference is of course based on a stereotypical description of their work. One could say that Kuhn describes *how it is*, while Popper writes *how it should be*, and they in their own ways dealt with questions on the relationship of knowledge and power, and on whether unity of knowledge is possible. The main difference would be whether working within a certain paradigm (if ever you

can avoid *not* working within one), within certain accepted boundaries, and with defining away certain uncertainties, is more beneficial to scientific *progress* than constantly attacking established knowledge.

Working with imperfect, questionable, or even false statements (or indicators in our case) may actually generate results (Popper, 2002: 72). Testing theory leads to results that are *adequate*, or that we *decide to accept* (Popper, 2002: 86). Sometimes, it gets forgotten that certain *leaps of faith* have been made, and indicators become accepted as *truths*. In some way the accepted methods and indicators to measure government become a defining paradigm, and we see the emergence of a ‘spontaneous pattern of deference’ (Fuller, 2006: 5). A *decision* is taken to *believe* something; in our case, that an indicator or part of an indicator is correct.

This deference, these beliefs, help to expand empirical and econometric analyses of the overall performance of government. But as a result, this *lock-in* makes rejecting the basic assumptions or leaps of faith increasingly difficult. We are thus faced with a dilemma: do we continue constantly undermining current efforts to measure government, or do we accept a number of working assumptions, knowing these assumptions can never be proved correct, and are indeed very likely to be incorrect? There is no solution for this dilemma. Even Popper and Kuhn disagreed on how much critique an established theory should get. This accepted level of critique appears to be discipline-specific, where political scientists revel in constantly undermining others’ efforts at theorising, and where economists feel much more comfortable working within a set of assumptions. Neither is a certain path to *progress*.

The choice we have to make appears to be one where one option will lead to endless discussions and no indicators whatsoever, and the other to a composite indicator that is both incorrect and of little use from an academic perspective (and some would argue also from a policy perspective). The second option is likely to prevail. Political science and public administration operates in a way that is quite far removed from the established scientific *ideal*, while the economic/econometric approach has a much higher appearance of rationality and coherence. Kuhn, in the preface to his *The structure of scientific revolutions* noted how he ‘was struck by the

number and extent of the overt disagreements between social scientists about the nature of legitimate scientific problems and methods' (Kuhn, 1962: x). Or, phrased differently,

[...] Kuhn identified his Eureka moment - when his theory of paradigms finally gelled-as occurring when he witnessed the vast difference in the way social and physical scientists conduct arguments. No matter how much physicists disagreed on the value of a particular piece of research, they could always agree on an exemplar against which to judge it. This was not possible in the social sciences, where any candidate exemplar (say Marx, Durkheim, Keynes Freud, Skinner, or nowadays, Foucault) would also be a lightning rod for fundamental disagreements (Fuller, 2006: 21-2).

Of the social science disciplines, economics is the one that has gone the furthest in adapting the natural science approach. It is not at all certain that this is the right way to go. Fuller criticised followers of Kuhn who interpreted Kuhn's work as a manual to turn their field into a full-fledged science (Fuller, 2006: 22). Kuhn, indeed, stated that 'both history and acquaintance made me doubt that practitioners of the natural sciences possess firmer or more permanent answers to such questions than their colleagues in social science' (Kuhn, 1962: x). Despite this remark, political science and public administration are at a disadvantage when it comes to expressing its views of how government should be measured -- or better, its views of how government cannot be measured unequivocally. Political science and public administration (with the exception of the game theorists and formal modellers in political science) lack the *appearance* of rationality and scientificness that a discipline such as econometrics has.

Can and should there be a scientific measurement of government?

Public administration, as a social science, is known for its non-paradigmatic character (Rommel and Christiaens, 2006; Flyvbjerg, 2001; Flyvbjerg, 2005: 40; Lan and Anders, 2000). There is no

set of 'universally recognised scientific achievements that for a time provide model problems and solutions to a community of practitioners' (Kuhn, 1962: x). There is no *normal science* in the public administration discipline. There have been calls for public administration to become more of a natural science, and the discipline has heard many complaints about its lack of empiricism (Cruise, 1997). If we want to be more like natural sciences, we need more generally agreed-upon definitions, which we are unlikely to get. By accepting a common definition of *government*, and by agreeing on government's objectives, the discipline would be able to use its energy for conducting more risky measurement and for building complicated (mathematical) explanatory models for explaining differences in government performance. Following such a path requires the acceptance of a key set of universal assumptions. Yet it is also a risky strategy. Once the core assumptions are challenged, the entire system collapses. The medieval discussion on the gender of angels only stopped making sense once their existence became to be doubted, and astrology appears to be more or less logically consistent when you accept its assumptions.

The calls for a more *scientific* public administration scholarship have lead researchers to focus on specific aspects of government and public sector management, rather than on more general questions of public administration. Mosher (in 1968) complained that students of government 'prefer to study those subjects which are amenable to scientific, objective, and quantifiable treatment. A declaration that these topics are somewhat less important than they seem would be self-defeating' (Mosher, 1968: 6). Dror likewise remarked that those following the methods of physical sciences in developing policy knowledge tend to focus on micro-issues, not macro (Dror, 1973). Fukuyama reacted against economists' tendency to apply their methodological individualism on traditional social science topics, looking for general laws. He made a case that the methods used by economists are unlikely to lead to a better understanding of the functioning of public administration: 'The effort to be more "scientific" than the underlying subject matter permits carries a real cost in blinding us to the real complexities of public administration as it is practiced in different societies' (Fukuyama, 2004: 123).

A traditional view of science is that it is a search for ‘a measure of the degree of law-likeness or regularity of events’ (Popper, 2002: 123). Such a desire for simplicity makes it quite convenient no longer to mention the leaps of faith or assumptions behind the analysis. The tests and models become technically very reliable, but at the same time quite detached from the *reality* they claim to measure. The search for laws and regularities is also built on a teleological assumption, which in the case of the study of government would mean there is a *final* form of government, that is both more advanced and better than the one we have now. There is an assumption that there is a *best* or ideally performing government. Analysing the performance of government and its determinants is then to contribute to an improvement of government. It is however disputed whether social theory is a cumulative process (Flyvbjerg, 2005).

Conclusion

The argument developed in this article is a fairly simple one: we cannot measure government because we cannot define it. To measure, you need firm ground (Stone, 2002: 65). Yet the juxtaposition of the two approaches to measuring government presented in the previous section is a somewhat caricatural one, presenting them as incompatible. In reality, we have seen quite a lot of cross-fertilisation and *rediscovering* in both disciplines.

What we are really dealing with is the challenge of *empiricising* the study of public administration, while at the same time *theorising* econometrics. From the side of public administration, this will have to imply a hard decision between going ahead with *good enough* measurement, or leaving this subfield (measuring government) to others while disagreeing with them. Are we willing to take certain decisions that will help progress in measurement and analysis, or do we refuse to make such arbitrary decisions? Econometricians, from their side, need to have more attention for the fact that measuring government performance is not just about techniques, and that more clarity is needed about the theoretical models behind the measurement, and about how the indicators in the models have been constructed.

The discussion on the direction of future efforts to measure and compare government is not just a philosophical one. It has important implications for how the results can be used. Are we striving

for academically *correct* measurement, or do we want to use measurement as a tool for improvement? We are faced with a difficult choice between a technically more correct measurement that may be useless and irrelevant for improvement, and a technically incorrect measurement that is useful for public sector improvement.

Perhaps we should not attempt measuring government in general, and focus instead on parts of the public sector that can be properly defined and/or that lend themselves to relevant measurement (see e.g. Wilson, 1989). After all, nobody would ever attempt measuring *the private sector* by using a single indicator, precisely because it is such a complex thing. Yet, given the novelty of this type of research, there is a first mover advantage: by constructing a composite indicator for government performance, you limit others' manoeuvring space to criticising it, to adapting it, or to working within this framework.

Measurement will always be context and time dependent, reflecting the current consensus, or majority view, of what government's objectives should be (also in the research community). Every outcome or objective that is used in models or measurement is to some extent a political choice. These political preferences can sometimes be much-contested (e.g. small state bias), but are in many other cases less problematic (e.g. life expectancy). Measurement can therefore only work within a well-defined time frame, and in a context where policy learning has created a working agreement on government objectives. Such situations are best visible where the definition of objectives is mediated by international organisations (Human Development Index), and where a leap of faith has been firmly established and is acceptable for most participants in the debate. Such an agreement does at present not exist about government objectives.

Reference List

Adams, G.B. and Balfour, D.L. (1998) *Unmasking Administrative Evil*, Thousand Oaks: Sage.

Afonso, A., Schuknecht, L. and Tanzi, V. (2003) 'Public Sector Efficiency: An International Comparison'. European Central Bank Working Paper no. 242.

Afonso, A., Schuknecht, L. and Tanzi, V. (2005) Public Sector Efficiency: an International Comparison, *Public Choice*.

123: 3-4 pp321-347.

Baumol, W. (1967) Macroeconomics of Unbalanced Growth: the Anatomy of the Urban Crisis, *American Economic Review*. 57 pp415-426.

Behn, R.D. (2003) Why Measure Performance? Different Purposes Require Different Measures, *Public Administration Review*. 63: 5 pp586-606.

Berlin, I. (1991) 'The Pursuit of the Ideal' in I. Berlin *The Crooked Timber of Humanity*. London: Fontana Press, pp1-19.

Bertok, J., Hall, J., Kraan, D.-J., Malinska, J., Manning, N., and Matthews, E. (2006), *Issues in Outcome Measurement for 'Government at a Glance'*. OECD Project on Management in Government Technical Paper 3, OECD, Paris.

Bevan, G. and Hood, C. (2006) What's Measured Is What Matters: Targets and Gaming in the English Public Health System, *Public Administration*. 84: 3 pp517-538.

Blondel, J. (1982) *The Organization of Governments: A Comparative Analysis of Governmental Structures*, London: Sage.

Boas, M. (2001) 'Multilateral Development Banks and Sustainable Development: The Strategy of Depoliticisation' in F. Biermann, R. Brohm K. Dingwerth, K. (eds) *Proceedings of the 2001 Berlin Conference on the Human Dimensions of Global Environmental Change "Global Environmental Change and the Nation State"*. Potsdam: Potsdam Institute for Climate Impact Research, pp434-440.

Bonar Blalock, A. (1999), Evaluation Research and the Performance Management Movement. From Estrangement to Useful Integration?, *Evaluation*. 5: 2 pp117-149.

Bovaird, T. (2005) Public Governance: Balancing Stakeholder Power in a Network Society, *International Review of Administrative Sciences*. 71: 2 pp217-228.

Bovaird, T. and Loeffler, E. (2003) Evaluating the Quality of Public Governance: Indicators, Models and Methodologies, *International Review of Administrative Sciences*. 69: 3 pp.313-328.

Boyle, D. (2001) *The Tyranny of Numbers: Why Counting Can't Make Us Happy*, London: Harper Collins.

Boyle, R. (2006) *Measuring Public Sector Productivity: Lessons From International Experience*, CPMR Discussion Paper No.35, Dublin: Institute of Public Administration.

International comparisons of public sector performance

- Bozeman, B. (2004) *All Organizations Are Public: Comparing Public and Private Organizations*, Washington D.C.: Beard Books.
- Brunsson, N. (1989) *The Organization of Hypocrisy: Talk, Decision and Actions in Organizations*, Chichester: John Wiley & Sons.
- Brusis, M. and Ramel, J. (2006) 'Reformfähigkeit messen: Konzeptionelle Überlegungen zu einem Reformfähigkeitsindex für OECD-Staaten'. Papier für das Panel der Sektion Staatslehre und politische Verwaltung, 23. DVPW-Kongress, Münster, 25-29 September.
- Caiden, G. (1968) Administrative Reform, *International Review of Administrative Sciences*. 34 pp347-358.
- Cherchye, L., Moesen, W., Rogge, N. and Van Puyenbroeck, T. (2007) An Introduction to 'Benefit of the Doubt' Composite Indicators, *Social Indicators Research*. 82: 1 pp111-145.
- Cirillo, S. (2006) 'Australia's Governance Aid: Evaluating Evolving Norms and Objectives'. Asia Pacific School of Economics and Government: Policy and Governance Discussion Paper 06-01.
- Cruise, P.L. (1997) Are Proverbs Really So Bad? Herbert Simon and the Logical Positivist Perspective in American Public Administration, *Journal of Management History*. 3: 4 pp342-359.
- Cyert, R.M. and March, J.G. (1963) *A Behavioral Theory of the Firm*, Englewood Cliffs, NJ: Prentice-Hall, Inc.
- De Bruijn, H. (2002) *Managing Performance in the Public Sector*, London: Routledge.
- Doornbos, M. (2003) "Good Governance": the Metamorphosis of a Policy Metaphor, *Journal of International Affairs*. 57: 1 pp3-17.
- Dror, Y. (1973) *Public Policymaking Reexamined*, Aylesbury: Leonard Hill Books.
- Eisenberg, E.M. (1984) Ambiguity As Strategy in Organizational Communication, *Communication Monographs*. 51 pp227-242.
- 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 October.
- Flyvbjerg, B. (2001) *Making Social Science Matter: Why Social Inquiry Fails and How It Can Succeed Again*, Cambridge: Cambridge University Press.
- Flyvbjerg, B. (2005) Social Science That Matters, *Foresight Europe*, October pp38-42.

International comparisons of public sector performance

- Fukuyama, F. (2004) *State Building: Governance and World Order in the Twenty-First Century*, London: Profile Books.
- Fuller, S. (2006) *Kuhn Vs. Popper: the Struggle for the Soul of Science*, Thriplow: Icon Books.
- Grindle, M.S. (2004) Good Enough Governance: Poverty Reduction and Reform in Developing Countries, *Governance: an International Journal of Policy, Administration and Institutions*. 17: 4 pp525-548.
- Handler, H., Koebel, B., Reiss, P. and Schratzenstaller, M. (2005) 'The Size and Performance of Public Sector Activities in Europe', WIFO Working Paper 246/2005. Vienna: Österreichisches Institut für Wirtschaftsforschung.
- Hewitt de Alcántara, C. (1998) Uses and Abuses of the Concept of Governance, *International Social Science Journal*. 50: 155 pp105-113.
- Kaufman, H. (1956) Emerging Conflicts in the Doctrines of Public Administration, *The American Political Science Review*. 50: 4 pp1057-1073.
- Kuhn, T.S. (1962) *The Structure of Scientific Revolutions*, Chicago: The University of Chicago Press.
- Kuhry, B. ed. (2004) *Public Sector Performance: An International Comparison of Education, Health Care, Law and Order and Public Administration in 29 Western Countries*, Den Haag: Sociaal en Cultureel Planbureau.
- Lan, Z. and Anders, K.K. (2000) A Paradigmatic View of Contemporary Public Administration Research: an Empirical Test, *Administration and Society*. 32: 2 pp138-165.
- Lindblom, C.E. (1980) *The Policy-Making Process*, Englewood Cliffs: Prentice-Hall.
- Lindblom, C.E. and Cohen, D.K. (1979) *Usable Knowledge: Social Science and Social Problem Solving*, New Haven and London: Yale University Press.
- March, J.G. ed. (1988) *Decisions and Organizations*, Oxford: Basil Blackwell Ltd.
- March, J.G. and Olsen, J.P. (1976) 'Organizational Choice Under Ambiguity', in J.G. March and J.P. Olsen (eds) *Ambiguity and Choice in Organizations*. Bergen: Universitetsforlaget, pp10-23.
- Martin, S. and Smith, P.C. (2006) Multiple Public Service Performance Indicators: Towards an Integrated Statistical Approach, *Journal of Public Administration Research and Theory*. 15: 4 pp599-613.
- Mosher, F.C. (1968) *Democracy and the Public Service*, New York: Oxford University Press.

International comparisons of public sector performance

- Murray, R. (1987) 'Productivity Measurement in Bureaucratic Organisations' in J.-E. Lane (ed.) *Bureaucracy and Public Choice*. London: Sage, pp171-194.
- O'Mahony, M. and Stevens, P. (2006) 'International Comparisons of Output and Productivity in Public Service Provision: A Review' in G.A. Boyne, K.J. Meier, L.J. O'Toole and R.M. Walker (eds) *Public Service Performance: Perspectives on Measurement and Management*. Cambridge: Cambridge University Press
- Oakeshott, M. (1991) 'Rationalism in Politics' in Oakeshott, M. *Rationalism in Politics and Other Essays*. Indianapolis: Liberty Fund, pp5-42.
- Olsen, J.P. (2004) Citizens, Public Administration and the Search for Theoretical Foundations, *PS*, January pp69-79.
- Olson, M.Jr., Sarna, N. and Swamy, A.V. (2000) Governance and Growth: A Simple Hypothesis Explaining Cross-Country Differences in Productivity Growth. *Public Choice*. 102: 3-4 pp341-364.
- Ostrom, V. (1974) *The Intellectual Crisis in American Public Administration* (revised edition), University, AL.: The University of Alabama Press.
- Pollitt, C. (2000) Is the Emperor in His Underwear? An Analysis of the Impact of Public Management Reforms, *Public Management Review*. 2: 2 pp181-199.
- Pollitt, C. (2003) *The Essential Public Manager*, Maidenhead: Open University Press.
- Popper, K. (2002) *The Logic of Scientific Discovery*, London: Routledge.
- Quinn, J.B. (1980) *Strategies for Change: Logical Incrementalism*, Homewood: Irwin.
- Radin, B.A. (2006) *Challenging the Performance Movement: Accountability, Complexity, and Democratic Values*, Washington, D.C.: Georgetown University Press
- Rommel, J. and Christiaens, J. (2006) Beyond the Paradigm Clashes in Public Administration, *Administrative Theory & Praxis*. 28: 4 pp610-617.
- Rosenbloom, D.H. (1983) Public Administration Theory and the Separation of Powers, *Public Administration Review*. 43: 3 pp219-227.
- Rothstein, B. and Teorell, J. (2005) 'What is Quality of Government? A Theory of Impartial Political Institutions', Paper at the 2005 Annual Meeting of the American Political Science Association, September 1-4.
- Rutgers, M.R. (2001) Traditional Flavors? The Different Sentiments in European and American Administrative

International comparisons of public sector performance

- Thought, *Administration and Society*. 33: 2 pp220-244.
- Saltelli, A. (2007) Composite Indicators Between Analysis and Advocacy, *Social Indicators Research*. 81: 1 pp65-77.
- Sanderson, I. (2002) Evaluation, Policy Learning and Evidence-Based Policy-Making, *Public Administration*. 80: 2 pp1-22.
- Santiso, C. (2001) International Co-Operation for Democracy and Good Governance: Moving Towards a Second Generation?, *European Journal of Development Research*. 13: 1 pp154-180.
- Simon, H.A. (1976) *Administrative Behavior: A Study of Decision-Making Processes in Administrative Organization* (3rd edition), New York: Free Press.
- Simon, H.A., Smithburg, D.W. and Thompson, V.A. (1959) *Public Administration*, New York: Alfred A. Knopf.
- Stone, D. (2002) *Policy Paradox: The Art of Political Decision Making* (revised edition), New York: W. W. Norton & Company.
- Van de Walle, S. (2006) The State of the World's Bureaucracies, *Journal of Comparative Policy Analysis*. 8: 4 pp439-450.
- Wilson, J.Q. (1989), *Bureaucracy: What Government Agencies Do and Why They Do It*, New York: Basic Books.