Performance management in Europe: An idea whose time has come - and gone?

WOUTER VAN DOOREN¹, CONNY HOFFMANN²

Full Reference:

Van Dooren, W., & Hoffmann, C. (2018). Performance Management in Europe: An Idea Whose Time Has Come and Gone?. In E. Ongaro & S. Van Thiel (Eds), *The Palgrave Handbook of Public Administration and Management in Europe* (pp. 207-225). Palgrave Macmillan, London.

ABSTRACT

New Public Management reforms in Europe, as elsewhere, heavily rely on performance indicators and targets. All corners of the public sector, from local to European and from policy formulation to management practice, have been affected. This focus on measurement fits well in a long tradition of measurement and state building. Yet, in recent years, disenchantment with performance management grows. More often than not, target regimes produce dysfunctional consequences. While the performance of performance target regimes is wanting, performance management is being reinvented. Rather than a system of accountability, performance management should prompt learning and dialogue. Performance management as a learning system may well be the next idea whose time has come.

¹ Wouter Van Dooren is Associate Professor at the research group 'Public Administration & Management', Department Political Science, University of Antwerp, Belgium.

Contact: wouter.vandooren@uantwerpen.be

² Conny Hoffmann is a doctoral Researcher at the research group 'Public Administration & Management', Department Political Science, University of Antwerp, Belgium.

Contact: cornelia.hoffmann@uantwerpen.be

CONTENTS

1	Introduction	3
2	The bedrock: modernisation and quantification	3
3	The heydays: an idea whose time has come	5
4	Performance management on the European level	7
5	The performance mind-set: assumption vs. reality	8
6	The disenchantment: an idea whose time has gone	9
7	The prospect: PM as a learning system - an idea whose time will come?	12
8	Conclusion	16
9	References	17

1 INTRODUCTION

This chapter reports on the rise, fall, and potential resurrection of performance management in Europe. The first chapter argues that in modern societies, performance measurement and management benefited from a strong tailwind. Our culture places great trust in numbers. Performance management therefore fits within a long lineage of quantification. The heyday of performance management was the *New Public Management* (NPM) era of the 1990s and 2000s. The political and societal context provided a fertile ground for integrating performance metrics in policy, management, and governance. The NPM edition of performance management purposively used indicators for steering behaviour, while previous efforts of quantification mainly aimed at understanding and control. The second section addresses the NPM version of performance management. In recent years, NPM-flavoured performance management systems were running into trouble, with still a strong tailwind but rough weather and high waves ahead. The attractiveness of indicators remained, but the performance of performance management systems was wanting. The impact of performance indicators is often absent or contrary to the expected outcome. We document this disenchantment with performance management in section 3. Some new avenues for performance management are however taking shape. Performance indicators should be primarily used for learning and improvement rather than for settling scores. We develop this argument in the final section.

2 THE BEDROCK: MODERNISATION AND QUANTIFICATION

Performance management is a modernisation project of which quantification is a central tenet (Porter, 1995). The appeal of performance management in the 1990s and 2000s can only be understood with reference to these bigger, long-term trends in society and government. The cultural fit between performance management and modernist visions of control explains why governments in the 1990s enthusiastically embraced performance management. The modernist appeal also explains the tenacity of performance management despite the mounting evidence of dysfunctionality. Performance measurement holds the promise of a society that is knowable and changeable through government intervention. Therefore, managers build scorecards with traffic lights mimicking a social engineer who controls society from a cockpit. This modernist legacy in contemporary management touches the roots of the building of the nation state.

In order to govern the world, administrators need to make the world legible (Desrosières, 1998; Scott, 1998). The natural world and real social patterns are too complex for administrative manipulation. Therefore, states attempt to reduce local complexity through systems of standardization (Brunsson, Rasche & Seidl, 2012; Pollitt, 2008; Timmermans & Epstein, 2010). Standardisation of measurement is one of the main battlefields in the fight for

central legibility. The development of the modern nation states and the ascendancy of centralised, uniform measurement systems go hand in hand. The development of the metric system in France is a good example of measurement and modernisation (Scott, 1998). Before modernity, units of measurement were based on local practices and requirements. A *pinte* in France, for instance, was equivalent to 0,93 litres or 3,33 litres depending on the region. Land yields in Ireland were expressed in terms of the number of cows a field could feed, which is the measure of interest for impoverished farmers (Scott, 1998). For the administrative state, these local measures were illegible and impractical. Napoleonic state builders hence enforced the metric system in the empire in order to facilitate the administration of France. Uniform measures would even contribute to equality of citizens under the law. Today, standardisation of measurement remains one of the major challenges of international trade and economic cooperation. Similarly, standards of human rights hold the promise of global citizenship.

The modernisation of society, including the attempt to make the world legible through measurement, is the bedrock for the development of the performance movement. Outcomes of public services, which later were re-labelled as performance, have been quantified throughout modern history (Van Dooren, 2008). Social reformers at the end of the 19th century used poverty maps for knowledge and advocacy. Charles Booth, for instance, drafted detailed poverty maps of London in order to confront government with its failings (Bulmer, Bales & Sklar, 1991). The progressive movement in the United States similarly advocated evidence-based policies (Stivers, 2000). At the beginning of the 20th century, the New York Bureau of Municipal Research employed most of the concepts of contemporary performance management, including outputs, outcomes, efficiency, and effectiveness (Williams, 2003, 2004). The bureau also developed the first ideas on performance budgeting. However, the ideas were not widely implemented. During the aftermath of World War II, national accounting and the measurement of economic performance were developed. In the 1960s, *Zero Based Budgeting* (ZBB) and the *Planning Programming Budgeting System* (PPBS) attempted to turn the budget into a performance management tool, albeit with limited success (Wildavsky, 1969). Management ideas such as PPBS and ZBB were exported from the United States to Europe. Meanwhile, an increasing number of societal sectors were subjected to measurement. The so-called *social indicator movement* started measuring in health, employment, education, and well-being (Bauer, 1966). In the 1970s, environmental performance received more attention in response to the first indications that economic growth had limits.

Performance management has a long lineage, but is essentially a project of modernity, of making complex local practices legible for science, administration, and policy. In the 1990s, performance measurement became fully integrated into the daily practice of public management.

3 THE HEYDAYS: AN IDEA WHOSE TIME HAS COME

Most measurement efforts so far addressed societal issues: mapping poverty, establishing economic performance, measuring unemployment and health, and calculating the state of the environment. Measurement provided knowledge for administrative development and policy proposals. In the 1990s, however, measurement took a fundamentally different turn (Van Dooren, Bouckaert & Halligan, 2015). Measures no longer just measured, but became indicators to purposively guide administrative behaviour. Knowledge creation was no longer the sole, or even primary, purpose of measurement. Numbers are not used primarily for understanding, but for steering the behaviour of actors. Performance measures evolved into performance controls and are used for accountability (Ferlie, 1996; Kettl, 2002). This is the time of performance contracts, performance incentives, and performance pay. In other words, performance measurement transformed into performance management. Instead of inputs, rules, or processes, results have to guide managerial intervention.

While the idea of performance management was not new, (we trace its origins to New York in the early 20th century), in the 1990s, its day had come. Politics in developed democracies shifted towards the centre of the political spectrum. After capitalism and communism – the great ideologies of the cold war – politicians attempted to formulate a new middle ground (Giddens, 1998). They found a compromise between left-wing and right-wing ideologies by having recourse to scientific evidence. These initiatives collectively became known as the evidence-based policy movement. Their catchphrase was what matters is what works. Not ideology but evidence should inform policy decisions. The reliance on science rather than ideology allowed for some uncommon coalition governments between left-wing and right-wing parties in the Netherlands, headed by Wim Kok, and in Belgium, headed by Guy Verhofstadt. Tony Blair's New Labour and Bill Clinton's Democratic Party moved to the centre and towards the median voter. Left-leaning leaders such as Gerhard Schröder in Germany and Massimo D'Alema in Italy also adopted the third way politics (Giddens, 2013). Centrist politics was not compatible with deep-seated ideological conflict. Evidence-based policy offered a way out of the trenches.

Third way politics also had to define a solution for the state and its bureaucracy. You cannot simultaneously have more state and less state. NPM offered a solution for this apparent deadlock. NPM promised more government and less government at the same time. Al Gore, vice president under Bill Clinton, pledged a government that performs better and costs less (Gore, 1993). Osborne and Gaebler's (1993) metaphor of a government that *should steer, not row* summarizes NPM accurately. The state was to be more actively involved in the public domain, but not with public services or with rules and regulations. The state was to set performance targets for various private actors and not to interfere with the attainment of results. This vision of a small state that regulates performance was attractive to both left and right, united in their distrust of traditional bureaucracy (Goodsell, 1994). NPM

hinged upon performance indicators (Bouckaert & Peters, 2002; Hood, 1991). The steering state used performance measurement as its compass. Performance indicators held actors accountable. Managers were free to manage, but also had to account for their results (Lægreid, 2000). Resource allocation and rewards were linked to performance targets. Instead of traditional input-based budgets, many countries implemented performance budgeting (Curristine, 2005; Pollitt & Bouckaert, 2011). Competition within the public sector was promoted, while performance indicators kept score. Rankings of schools, hospitals, and many other public services were intended to trigger customer choice and quasi-market pressures (Bevan & Hood, 2006; Jilke, 2015). Hierarchical relations within the public sector were dissolved and replaced by contractual relationships, including performance targets (Greve, Flinders & Van Thiel, 1999; Verhoest, 2005). Managerial practice within the public sector was to mimic private management (Pollitt & Bouckaert, 2011). For many, private management equalled hard-nosed, numbers-based management. The success of Kaplan and Norton's (1996) balanced scorecard in the public sector resulted from this private sector mimicry. The scorecard required the identification of key performance indicators that would elicit attention for core tasks, similar to the way in which private companies focus on the bottom line. Finally, privatisation of public services coincided with new forms of contracting with private providers (Greve, 2007). The performance approach demanded that contracts specify outputs rather than inputs or procedures. Contracts would no longer demand the building of a road, the delivery of a bus, or the provision of a job placement programme. Rather, contracts specified availability of infrastructure, the frequency of a bus line, or the number of unemployed finding a job.

Despite the globally uniform NPM story, European governments implemented performance management in vastly different ways (Christensen, Lægreid & Wise, 2002; Pollitt & Bouckaert, 2011; Premfors, 1998). The United Kingdom was the early adopter of the performance management doctrine in Europe. In the 1980s, Thatcherite reforms focussed on privatisation and reducing the size of the public sector. The strategy was mainly to minimize the public sector (Pollitt & Bouckaert 2011). In the late 1990s, however, the emphasis shifted from privatisation to changing management practice within government. In terms of Pollitt and Bouckaert's trajectories, the shift went from minimizing to modernising. The British performance management framework became highly detailed, with a top down design, targets, service standards, and performance assessments.

Continental Europe adopted performance management regimes more hesitantly. Reforms were transformed to suit the local practice (Christensen & Lægreid, 2007). In France, the main performance management initiative was a performance budgeting scheme. Administrative reform was, however, embedded in a larger reform of the state that led to decentralisation and deconcentration of central tasks to local governments and independent agencies respectively (Kickert, 2011). Germany was a modest adopter of NPM. German unification required a rapid integration of Eastern bureaucracies into the West-German administrative fabric. There was no time to concep-

tually revise the Weberian administrative system (Kickert, 2011). Moreover, the legal administrative culture was barren soil for performance management. The majority of performance management initiatives in Germany existed at local and city level. Italy also adopted several performance management schemes (Ongaro, 2009). Performance management was, however, not imposed throughout government. Rather, performance management resulted from sectorial legal initiatives and performance funding, particularly in health care and education. Similar to the situation in Germany, however, performance management was primarily a bottom up process, starting with local governments and followed by some regional and central entities.

4 PERFORMANCE MANAGEMENT ON THE EUROPEAN LEVEL

Performance indicator schemes were not only found on national or sub-national levels. From the 1990s onwards, performance indicators featured in European governance. In the 1990s, economic governance, including the accession to the Euro zone, centred on the Maastricht criteria, which were five economic indicators of economic performance. The public deficit norm of 3% of the GDP and the public debt norm of 60% of the GDP were the most influential targets. Member states actively worked towards meeting the targets, partly by cleaning up the budget and partly by juggling the statistics, as was later discovered (Reich, 2015).

The apparent success of the Maastricht criteria in steering national policies translated in the official governance policy of the *Open Method of Coordination (OMC)* (European Commission 2001a). This method had five components:

- agreement on common objectives;
- establishing common indicators as a means of comparing best practice and measuring progress;
- translating the EU objectives into national/regional policies through the development of National Action Plans;
- publishing reports and analysing and assessing the National Action Plans;
- establishing a Community Action Program to promote policy cooperation and transnational exchange of learning and good practice.

Indicators and targets played a prominent role in this process. In line with NPM, the main purpose was to steer the behaviour of member states and not to obtain knowledge for policy.

The OMC was applied in the Lisbon strategy (2000-2010). At the Lisbon European Council in spring 2000, the European Union set the “*strategic goal for the next decade: to become the most competitive and dynamic knowledge-based economy in the world capable of sustainable economic growth with more and better jobs and greater social cohesion*” (Paragraph 5 of the Council conclusions)

(European Commission, 2000). Each year, member states had to account for their progress as measured by a set of indicators. These progress reports were intended to push countries in the right direction. The indicator set commenced with 35 indicators that addressed employment, innovation, economic reform, and social cohesion. Soon, the set expanded to 42 indicators that also covered environment and some general economic indicators. Besides the selection of structural, key indicators, hundreds of other indicators were collected in complementary processes. Unlike the Maastricht process, the Lisbon strategy was generally viewed as a failure. In a mid-term review, Wim Kok attributed the slow start of the Lisbon strategy to an excess of indicators that eroded their steering capacity. The mid-term review report states that: *'More than a hundred indicators have been associated with the Lisbon process, which makes it likely that every country will be ranked as best at one indicator or another. This makes the instrument ineffective. (...) The European Commission should present the Heads of State or Government and the wider public annual updates on these key 14 Lisbon indicators in the format of league tables with rankings (1-25), praising good performance, and castigating bad performance – naming, shaming, and faming.'* (European Commission, 2004, p. 43)

It clearly did not occur to the authors of the report that countries may be better at one indicator or another, but not at all indicators at the same time. The midterm review mentions that indicator-based accountability is incompatible in the presence of many indicators. The fiscal crisis marked a return to economic governance. With the EU 2020 strategy, governance by indicators, however, persisted. The European surveillance systems were beefed up to cope with gaming strategies of countries (Buti & Carnot, 2012). Indicator-based governance was also implemented in other sectors of EU policy making, most notably with the stress tests of the European Banking Authority that monitor the sustainability of the European banks. In environmental policies, in 2013 the European Union adopted the Clean Air Quality package with an indicator-based governance scheme. The package included, inter alia, targets and air quality objectives for the period up to 2030. A revised *National Emission Ceilings Directive* determines stricter national emission ceilings for six main pollutants (i.e. six indicators). Breaching the targets can lead to legal proceedings with the Court of Justice in Luxembourg.

5 THE PERFORMANCE MIND-SET: ASSUMPTION VS. REALITY

Indicator-based governance has permeated all areas of governance. The engineering logic of performance management, with indicators and targets used as switches and controls, defined the performance management systems of the 1990s and the 2000s. Radin (2000, 2006) succinctly summarises and challenges seven major assumptions of what she calls the performance mind-set (Table 1).

Increasingly, disillusionment with the performance of performance management has crept into the ranks of scholars and policy makers alike.

Table 1: The assumptions of indicator-based governance

Assumption	Reality
Goals can be defined clearly and set firmly as the basis for performance measurement	Goals are volatile and difficult to measure
Goals are specific and the responsibility of definable actors	Goals are vague and responsibilities are shared
Outcomes can be specified independently of inputs, processes and outputs	There are feedback loops and back channels
Outcomes can be quantified and measured	Outcomes are often elusive
Outcomes are controllable and susceptible to external timing	Outcomes are unpredictable
Data is available, clear, and accurate	Data is often unavailable, unclear, and inaccurate
Results of the performance measurement can be delivered to an actor with authority to respond to the results	Political contexts seldom allow for direct action

Source: Radin (2000, 2006)

6 THE DISENCHANTMENT: AN IDEA WHOSE TIME HAS GONE

There is, recently, increasing criticism about performance management systems. We review this criticism using Hirschman's (1991) three theses of social interventions. Hirschman discusses the futility, jeopardy, and the perversity theses.

The *perversity* thesis implies that "any purposive action to improve some feature of the political, social, or economic order only serves to exacerbate the condition one wishes to remedy. The *futility* thesis holds that attempts at social transformation will be unavailing, that they will simply fail to 'make a dent'. Finally, the *jeopardy* thesis argues that the cost of the proposed change or reform is too high as it endangers some previous, precious accomplishment" (Hirschman, 1991, p. 7).

Translated to the criticism of performance management, perversity implies that performance outputs are manipulated. The behaviour that the performance indicator was intended to stimulate is not visible. On the contrary, actors display gaming behaviour that undermines the objectives of the performance management system. The futility thesis addresses the non-use of performance information. Information is collected, but it is not analysed or used. Finally, the jeopardy thesis describes juggling statistics, i.e. the manipulation of measurement. According to this thesis, resources invested in a measurement system could have better been invested elsewhere. Let us look at the support for these three theses in turn.

First, the *futility* thesis. Several authors refer to the "non-use" of performance information. Van Dooren, Bouckaert, and Halligan (2015) describe the various reasons why performance information might not be used. These include insufficient quality of the information, psychological barriers, cultural barriers, and the institutional context. Moynihan (2009) describes non-

use as “passive” use. Data is collected and disseminated, but not used. While this might initially appear to be of little consequence, closer examination reveals that the non-use of data is problematic. Particularly with regard to the ever-increasing demand for performance information, an increasing amount of information is collected, which requires substantial resources (time, money, and human resources). However, if the ‘data collectors’ see that their data is not used at all, this can lead to dissatisfaction and lower quality, as data collection is perceived as a ‘ticking box’ exercise. If data collectors provide data, the superiors will be satisfied. As nothing is subsequently done with the data, the data quality does not really matter, either. Hence, non-use of the data can negatively affect employees’ behaviour and motivation to deliver good performance.

Top down, comprehensive performance management system are more often than not ‘paper tigers’ (De Jong, 2016). A telling example is the comprehensive performance budgeting effort in the early 2000s in the Netherlands (De Jong, van Beek & Posthumus, 2013). The objective was to have all programmes in the budget covered by indicators or to explain why measurement would not be meaningful. The performance budget had its own performance target. For each department, a coverage rate of departmental programmes was calculated. Soon, the political debate turned towards the coverage rate instead of debating the substantive indicators on programme performance. People started wondering why departments had a low coverage rate rather than whether programmes were delivered efficiently and effectively. Did the department have something to hide? Were they not good managers? In response to public scrutiny, coverage rates went up. However, the indicators were not always very meaningful. The Ministry of Foreign Affairs included an indicator on the position of, inter alia, Afghanistan on the failed states index, the Health Ministry had an indicator on the number of medals won at the Olympics for a sports programme, and the Ministry of Education measured the number of ‘motivated’ students in higher education. The Netherlands now follow a different strategy for performance budgeting, with fewer indicators and a renewed emphasis on policy evaluation.

Second, the *jeopardy* thesis. Several authors describe incidents of measurement manipulation. Measurement manipulation can occur in different forms. The most common include indicator pollution, indicator inflation, misrepresentation, and misinterpretation (Bouckaert & Balk, 1991). Indicator pollution refers to the fact that input, output, and outcome indicators are mixed and undifferentiated. As a result, it is difficult to allocate responsibility. Indicator inflation describes the often-needless abundance of indicators. The addition of an increasing number of indicators to the performance management system does not aim to increase validity of the performance measurement, but to obfuscate analysis and decision-making. At the same time, the costs of measurement increase. Misrepresentation refers to reported behaviour that is different from actual behaviour, most notably in creative reporting and fraud. EU member states that invest off-budget through Public Private Partnerships, for instance, misrepresent the actual fiscal distress of a country. The euphemism of creative accounting reflects this practice. Mis-

interpretation is considered a result of the complexity of the issues at stake as well as the bounded rationality of the controllers (Smith, 1995). Whether intentional or not, misrepresented results jeopardize the operations of the organisation.

Policing provides us with many examples of gaming with indicators (Bernstein & Isackson, 2014; Wittebrood & Junger, 2002). The classification of offenses requires substantial interpretation. Offenses, for instance, can often be classified under different headings. An addicted, homeless, illegal immigrant who snatches a purse and jaywalks while running from the police has committed multiple violations. Depending on the targets and priorities of the police force, officers may register theft, drug abuse, rough sleeping, having no residence permit, jaywalking, or running from the police. On the other hand, they may decide to look the other way because other priorities are deemed more important. In all but the last case, police work is not fundamentally altered. Officers patrol the street and people are arrested. The registration, however, is tweaked towards the priority, which may jeopardize future policies and political debate. Bohte and Meier (2000) report a case of misrepresentation of public school performance in Texas. Students with expected low pass rates were deliberately excluded from certain exams to prevent lower overall results. While those students may receive decent education, they are excluded from the measurement (i.e. the test) in order to improve school performance. This gaming behaviour has to be understood in the context of the link between student results and teacher salaries, and the ratings of schools and teachers. As recent as 2015, another such case became public. In Atlanta, the “largest cheating scandal in the nation’s history” received much media attention as 11 teachers were brought before the court (and sentenced to prison) for having manipulated standardised tests by distorting the test scores (Blinder, 2015). Courty and Marschke (2007) found that within the scope of a job-training programme in the US, agencies had to terminate enrolment of their participants if they had not received any training services for 90 days. However, to improve agency performance and employment rate outcome, agencies decided to terminate enrolment once participants became employed, and not when training was completed.

Third, the *perversity* thesis. Recently, an increasing number of incidents have become known about how outputs are manipulated to give the impression of higher performance results. Manipulations of output include, for example, tunnel vision, which means focussing on quantifiable targets at the expense of unquantifiable targets (Smith, 1995). This includes, for example, teaching to the test in public schools to ensure that pupils have adequate results, which in turn, are linked to teacher salaries or school reputation (Grizzle, 2002; Wiggins & Tymms, 2002). Competences that are not easy to test are neglected in the classroom, such as attitudes, citizenship, or the skill of critical thinking. Schools with unprivileged children who do not acquire these competences at home should be able to focus on attitudes and values. Those schools feel the pressure of placing in the lower ranks of the league table, and hence will feel more pressure to teach the test. Another example of perversity is “cream skimming”. Cream skimming describes efforts to in-

crease performance results by deliberately choosing a “target group” that is likely to contribute positively to the performance, or deliberately exempting a certain target group that is likely to decrease performance results (Grizzle, 2002). The effect of the performance management system is the exclusion of precisely those target groups that would benefit the most from public services.

A shockwave went through the UK National Health Service (NHS) system when the Stafford hospital scandal was uncovered (Mid Staffordshire NHS Foundation Trust Public Inquiry, 2013). In 2008, the UK healthcare commission was alerted to the high mortality rates in patients admitted as emergencies. The enquiry that followed exposed serious deficiencies in the hospital’s management, which led to appalling conditions and patient neglect. While the full investigation revealed many shortcomings, part of what went wrong was attributed to the target regime of performance indicators in the hospital. Stafford was not an isolated case (House of Commons Public Administration Select Committee, 2003). Bevan and Hood (2006) analyse gaming in the hospital sector in the UK. They find considerable “creative reporting” regarding, for example, the waiting times in hospitals. Furthermore, in the Netherlands, the Dutch Forensic Institute has decreased waiting times by changing the definition of ‘waiting time’ (Frederik, 2016). Researchers at the institute were allowed to stop the timing process when they had to wait for further information from the police and the Public Prosecution Service (OM). On paper, this decreased the waiting times substantially, while in reality nothing had changed. Apparently, a few years earlier the waiting times had already been manipulated by placing quotas on the number of research requests to be received by the Forensic Institute within a month from the police and the OM. Hence, waiting times were reduced in the Forensic Institute, while the waiting lists at the police and the OM increased (where they were not included in the waiting time) (Frederik, 2016).

Performance management has not yet lived up to its aspirations. Rather than increasing efficiency and effectiveness, the premise under which performance management has been introduced, by measuring performance we actually might create the opposite effect (Hoffmann, 2016; Van Thiel & Leeuw, 2002). Van Thiel and Leeuw (2002) describe this effect as “performance paradox”, which refers to the “weak correlation between performance indicators and the performance itself”. As such, it deals with the reporting on performance, as opposed to performance itself (Van Thiel & Leeuw, 2002, p. 271).

7 THE PROSPECT: PM AS A LEARNING SYSTEM – AN IDEA WHOSE TIME WILL COME?

Against the backdrop of this disenchantment, calls for a change of direction have become louder. It is widely recognised that the occurrence of negative or perverse effects of performance information is related to the “hard use” of performance information, which refers to instances in which perfor-

mance information is tightly coupled to judgement (Van Dooren et al., 2015). This is particularly the case when performance information is used for accountability purposes (Olejniczak & Newcomer, 2014; Van Thiel & Leeuw, 2002). Consequently, an increasing number of scholarly contributions determine to find alternative ways of utilising performance information and holding organisations to account. One such way is to use performance information for learning purposes. To make such a system work, performance information would have to be decoupled from accountability mechanisms. In turn, organisations would not be held accountable for their performance results, but rather for the processes and learning efforts in which they are involved. This would furthermore imply a shift in the organisational culture. Creating a learning culture means that employees would not be afraid of committing “errors”. Rather, errors and failures would be regarded as learning opportunities, and not as a basis on which to judge employee performance (Olejniczak & Newcomer, 2014).

Two main forms of learning take place: single-loop and double-loop learning (Argyris & Schön, 1978). In the context of performance management, single-loop learning describes efforts to improve the existing performance measurement system by asking how existing routines could be done differently. Double-loop learning goes a step further and critiques the measurement system itself, by finding ways to do things differently (Common, 2004; Easterby-Smith, Crossan & Nicolini, 2000; Gilson, Dunleavy & Tinkler, 2009; Moynihan, 2005, 2011; Moynihan, 2008; Rogers & Williams, 2006; Van Dooren, 2011). Thus, whereas single-loop learning focusses on detecting and correcting errors, double-loop learning questions also the underlying norms, policies, and objectives of an organisation. Single-loop learning asks whether measurement is carried out correctly; double-loop learning asks whether the correct thing is measured. Hence, it is not surprising that organisations focus on single-loop learning. Furthermore, often organisations deliberately choose systems that make double-loop learning difficult (Argyris & Schön, 1978).

Double-loop learning in the context of performance management has important implications for the concept of performance management. Van Dooren et al. (2015) suggest that performance management should become more agile, political, and decentralised in order to serve the needs it has been created for in the first place (e.g. helping policy-makers make better informed decisions). This, however, runs counter to the prevailing concept of performance management as an “objective” and unambiguous system based on performance indicators for which actors are held accountable. In fact, performance management has to be recognised as ambiguous and dealt with as such. Including information from various sources and providing room for discussion and interpretation are examples of how to make performance management ambiguity-proof. This also implies that performance management must be made more agile by using performance indicators for learning purposes. Used in this way, performance indicators can be adjusted to the needs and contexts at stake – something that indicators for accountability generally cannot offer. We will go more into detail about this below. This

leads to the next point, decentralisation. Van Dooren et al. (2015) recommend that rather than building rigid, large, top-down systems, which are unlikely to provide timely information when needed, more power should be attributed to middle managers and frontline supervisors who have more knowledge of and experience with the situations at hand. Finally yet importantly, the political aspect of performance management should be acknowledged. Related to the first point of ambiguity, by engaging in a performance dialogue with different stakeholders, information can be better utilised, decisions more evidence-based, and the focus shifted towards outputs and outcomes.

Against the background of futility, perversity, and jeopardy of performance management, and the concepts of agile, decentral, and political performance management, Moynihan (2005; 2008) and Van Dooren et al. (2015) stress the importance of room for interpretation of performance information. Rather than using performance information in a “hard way”, and directly linking it to judgement or sanctions, it should be used in a “soft way”, by providing space for analysis and interpretation of the data before conclusions are drawn. In this context, Moynihan (2005; 2008) developed the interactive dialogue model, with “learning forums” as central. Learning forums are “routines that encourage actors to closely examine information, consider its significance, and decide how it will affect future action” (Moynihan, 2005, p. 205). Learning forums are thus a combination of structural and cultural elements: Structure, as learning forums require some form of “routine”, facilitation, and ground rules; Culture, as they centre on a non-confrontational approach, collegiality, equality, and dialogue. Both quantitative and qualitative knowledge are to be discussed in these forums. Hence, in these forums, performance information is analysed and interpreted, and thus used in a soft way, which is generally not sought after in strict accountability systems. Furthermore, once performance information is used for accountability purposes, it can no longer be used for learning purposes. There is no room for dialogue or interpretation, and actors adjust their behaviour to the needs or expectations of the account-holders, as the threshold for gaming or cheating is rather low. Thus, in many cases the informational content has little value. A recent study analyses the use of performance information in accountability- and learning-contexts, and finds that, even in a “most likely” case for learning (Finland), gaming or cheating take place (Hoffmann, 2016). In Finland, although much has been done to create a learning culture with different learning forums, pilot projects, and training programmes, apparently the accountability elements present in the system, most notably in form of performance contracts, partly “overrule” these learning efforts.

The figure below demonstrates the described mechanism. The more accountability prevails, the more dysfunctional effects (what we previously described as futility, jeopardy, and perversity) are going to occur. The relationship between accountability and dysfunctional effects is usually thought of as being linear (the solid line in the figure). However, literature and empirical evidence suggest that this relationship might not be linear, but resembles a curve (dashed line in figure 1) (Hoffmann, 2016). This has important

implications. It implies that even minimal accountability is detrimental for learning efforts. If performance information is intended for accountability purposes, it can no longer be used for learning purposes. The threshold for gaming and cheating to occur in accountability systems is low and, consequently, account-givers might already adjust their behaviour (or their data) to avoid blame or other negative consequences (Hoffmann, 2016).

Figure 1: Trade-off between accountability and learning (Hoffmann, 2016)

The idea that negative effects of targets can be mitigated by learning systems has previously been formulated (de Bruijn, 2002; Moynihan, 2008; Radin, 2006). Yet, the recommendation of most performance scholars thus far was to seek a middle ground between learning and accountability. An optimal balance of learning and target-based pressure should yield the best results. However, the incompatibility of learning and accountability (figure 1) suggests that it is difficult to have the best of both worlds. Hence, a choice exists between performance measurement for learning and performance measurement for accountability.

8 CONCLUSION

Performance management fits in a long legacy of modernisation. In order to build a central state apparatus that could be in control, local complexity needed to be made legible. Performance management and performance indicators held the promise of control and hence could be seen as an instance of high modernism (Van de Walle & Roberts, 2008). Performance management in NPM, however, also departed from established measurement traditions. The purpose of measurement was no longer understanding, legibility, or control. The purpose was to actively steer behaviour through performance contracts and rankings. NPM catchphrases capture this approach tersely: *what gets measured, gets done, only results matter, and governments should steer, not row*. Furthermore, steering must primarily be achieved with targets, not with rules and regulations. In recent years, however, disillusionment with the performance of performance management has grown. Many indicators are not used at all. The registration of indicators is subject to gaming. Other indicators lead to adverse effects. The evidence for the occurrence of these dysfunctional effects is accumulating and is by no means anecdotal (Hoffmann, 2016).

With increasing evidence of its shortcomings, performance management finds itself at a crossroads. The engineer's logic – set targets, measure attainment, and punish or reward – has reached its limits. A learning logic presents itself as a promising alternative. Instead of performance targets, we could have a performance dialogue (Moynihan, 2008). Nevertheless, a learning system does not come without its own drawbacks, either (Grieves, 2008; Lewis & Triantafillou, 2012). Organisational structures and cultures must be revisited. Control mechanisms would need to be replaced by trust mechanisms. Moreover, the true purpose of and need for performance information would have to be uncovered to avoid engaging in performance dialogues for the sake of simply doing so, rather than to envision genuine change and improvement.

9 REFERENCES

- C. Argyris and D.A. Schön (1978) *Organizational learning*. Reading (Mass: Addison-Wesley Pub. Co).
- R.A. Bauer (1966). *Social Indicators* (Cambridge: MIT Press).
- D. Bernstein and N. Isackson (2014) The Truth About Chicago's Crime Rates: Part 2. *Chicago magazine*. Consulted from: <http://www.chicagomag.com/Chicago-Magazine/June-2014/Chicago-crime-statistics/>, date accessed 19 May 2016.
- G. Bevan and C. Hood (2006) 'What's measured is what matters: Targets and gaming in the English public health care system', *Public Administration*, 84, 517–538.
- A. Blinder (2015) Atlanta educators convicted in school cheating scandal. *The New York Times*. New York.
- J. Bohte and K.J. Meier (2000) 'Goal displacement: Assessing the motivation for organizational cheating', *Public Administration Review*, 60, 173–182.
- G. Bouckaert and W. Balk (1991) 'Public productivity measurement: diseases and cures', *Public Productivity & Management Review*, 15, 229–235.
- G. Bouckaert and B.G. Peters (2002) 'Performance Measurement and Management: The Achilles' Heel in Administrative Modernization', *Public Performance & Management Review*, 25, 359–362.
- N. Brunsson, A. Rasche and D. Seidl (2012) 'The Dynamics of Standardization: Three Perspectives on Standards in Organization Studies', *Organization Studies*, 3, 613–632.
- M. Bulmer, K. Bales and K.K. Sklar (1991) *The Social Survey in Historical Perspective, 1880-1940* (Cambridge: Cambridge University Press).
- M. Buti and N. Carnot (2012) 'The EMU Debt Crisis: Early Lessons and Reforms', *JCMS: Journal of Common Market Studies*, 50, 899–911.
- T. Christensen and P. Lægreid (2007) *Transcending new public management: the transformation of public sector reforms* (Aldershot, England; Burlington, VT: Ashgate).
- T. Christensen, P. Lægreid and L.R. Wise (2002) 'Transforming Administrative Policy'. *Public Administration*, 80, 153–178.
- R. Common (2004). Organisational learning in a political environment. *Policy Studies*, 25(1), 35–49.
- P. Courty and G. Marschke (2007) 'Making government accountable: Lessons from a federal job training program', *Public Administration Review*, 65, 904–916.

- T. Curristine (2005) 'Performance Information in the Budget Process: Results of OECD 2005 Questionnaire', *OECD Journal on Budgeting*, 5, 54.
- H. De Bruijn (2002) 'Performance measurement in the public sector: Strategies to cope with the risks of performance measurement', *International Journal of Public Sector Management*, 15.
- M. De Jong (2016) *Why Agencies Budget For Results*. Erasmus University Rotterdam. Consulted from: <http://hdl.handle.net/1765/79334>
- M. De Jong, M. I. Van Beek and R. Posthumus (2013) 'Introducing accountable budgeting: Lessons from a decade of performance-based budgeting in the Netherlands', *OECD Journal on Budgeting*, 12, 1–34.
- A. Desrosières (1998) *The politics of large numbers: A history of statistical reasoning*. (Cambridge: Harvard University Press).
- M. Easterby-Smith, M. Crossan and D. Nicolini (2000). Organizational learning: Debates past, present and future. *Journal of Management Studies*, 37(6), 783-796.
- European Commission (2000) *The Lisbon European Council - An agenda of economic and social renewal for Europe: Contribution of the European Commission to the special European Council in Lisbon, 23-24th March 2000*. Brussels: European Commission.
- European Commission (2004) *Facing the challenge: the Lisbon strategy for growth and employment* (Text). Brussels: European Commission.
- E. Ferlie (1996) *The New Public Management in Action* (Oxford: Oxford University Press).
- J. Frederik (2016) Wachttijden verkorten? Verander gewoon de definitie even. *De Correspondent*.
- A. Giddens (1998) *The Third Way: the Renewal of Social Democracy* (Cambridge: Polity Press).
- A. Giddens (2013) *The Third Way and its Critics* (London: John Wiley & Sons).
- C. Gilson, P. Dunleavy and J. Trinkler (2009). *Organizational learning in government sector organizations: Literature review*. London: LSE Public Policy Group (PPG).
- C.T. Goodsell (1994) *The case for bureaucracy* (London: Chatham House Publishers).
- A. Gore (1993) *From Red Tape to Results: Creating a Government That Works Better & Costs Less. Report of the National Performance Review* (Washington DC: US Government Printing Office).
- C. Greve (2007) *Contracting for Public Services* (Routledge).

- C. Greve, M. Flinders and S. Van Thiel (1999) 'Quangos—What's in a name? Defining Quangos from a comparative perspective', *Governance: An International Journal of Policy and Administration*, 12, 129–146.
- J. Grieves (2008) 'Why we should abandon the idea of the learning organization', *The Learning Organization*, 15, 463–473.
- G.A. Grizzle (2002) 'Performance measurement and dysfunction: The dark side of quantifying work', *Public Performance & Management Review*, 25 363–369.
- A.O. Hirschman (1991) *The Rhetoric of Reaction: Perversity, Futility, Jeopardy* (Cambridge, Mass: Harvard University Press).
- C. Hoffmann (2016) *At a crossroads - How to change ways towards more meaningful performance management?* (Antwerp: University of Antwerp).
- C. Hood (1991) 'A public management for all seasons', *Public Administration*, 69, 3–19.
- House of Commons Public Administration Select Committee (2003) *On Target? Government by Measurement* (London: The Stationary Office).
- S. Jilke (2015) 'Choice and Equality: Are Vulnerable Citizens Worse Off After Liberalization Reforms?', *Public Administration*, 93, 68–85.
- R.S. Kaplan and D.P. Norton (1996) *The Balanced Scorecard: Translating Strategy into Action* Harvard Business School Press (Boston, Mass: Harvard Business School Press).
- D.F. Kettl (2002) *The Transformation of Governance* (Baltimore: John Hopkins University Press).
- W.J. Kickert (2011) 'Public management reform in continental europe: national distinctiveness', in *The Ashgate Research Companion to New Public Management* (Farnham: Ashgate), 97–112.
- P. Lægreid (2000) 'Top Civil Servants Under Contract', *Public Administration*, 78, 879–896.
- J.M. Lewis and P. Triantafillou (2012) 'From performance measurement to learning: A new source of government overload?' *International Review of Administrative Sciences*, 78, 597–614.
- Mid Staffordshire NHS Foundation Trust Public Inquiry. (2013). *Report of the Mid Staffordshire NHS Foundation Trust Public Inquiry: Executive Summary* (The Stationery Office).
- D. Moynihan (2005) 'Goal-based learning and the future of performance management', *Public Administration Review*, 65, 203–216.
- D.P. Moynihan (2008) *The dynamics of performance management: Constructing information and reform* (Washington, D.C: Georgetown University Press).

- D.P. Moynihan (2009) 'Through a glass, darkly: Understanding the effects of performance regimes', *Public Performance & Management Review*, 32, 592–603.
- K. Olejniczak and K. Newcomer (2014) Moving towards accountability for learning in K. Olejniczak and S. Mazur (Red.), *Organizational Learning. A Framework for Public Administration* (Warsaw: Scholar Publishing House), 81–98.
- E. Ongaro (2009) *Public Management Reform and Modernization: Trajectories of Administrative Change in Italy, France, Greece, Portugal and Spain* (London: Edward Elgar Publishing).
- D. Osborne and T. Gaebler (1993) *Reinventing Government: How the Entrepreneurial Spirit is Transforming the Public Sector from Schoolhouse to State House and City Hall to the Pentagon* (MA: Addison Wesley).
- C. Pollitt (2008) *Time, policy, management: governing with the past* (Oxford: Oxford University Press).
- C. Pollitt and G. Bouckaert (2011) *Public Management Reform: A comparative analysis-new public management, governance, and the Neo-Weberian state* (Oxford: Oxford University Press).
- T.M. Porter (1995) *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*. (Pinceton: Princeton University Press).
- R. Premfors (1998) 'Reshaping the democratic state: Swedish experiences in international perspective', *Public Administration*, 76, 141–159.
- B.A. Radin (2000) 'The government performance and results act and the tradition of federal management reform: Square pegs in round holes', *Journal of Public Administration Research and Theory*, 10, 111–135.
- B.A. Radin (2006) *Challenging the Performance Movement: Accountability, Complexity, and Democratic Values* (Washington DC: Georgetown University Press).
- R. Reich (2015). How Goldman Sachs Profited From the Greek Debt Crisis. *The Nation*.
Consulted from: <http://www.thenation.com/article/goldmans-greek-gambit/>
- P.J. Rogers and B. Williams (2006). Evaluation for practice improvement and organisational learning. In I. Shaw, J.C. Greene, and M.M. Mark (Eds), *handboek of evaluation: Policies, programs and practices* (pp. 76–97). London, UK: SAGE.
- J.C. Scott (1998) *Seeing like a state: How certain schemes to improve the human condition have failed* (New Haven: Yale University Press).
- P. Smith (1995) 'On the unintended consequences of publishing performance data in the public sector', *International Journal of Public Administration*, 18, 277–310.

- C. Stivers (2000) *Bureau Men, Settlement Women: Constructing Public Administration in the Progressive Era* (Kansas: University Press of Kansas).
- S. Timmermans and S. Epstein (2010) 'A World of Standards but not a Standard World: Toward a Sociology of Standards and Standardization', *Annual Review of Sociology*, 36, 69–89.
- S. Van de Walle and A. Roberts (2008) 'Publishing performance information: an illusion of control?' in W. Van Dooren and S. Van de Walle: *Performance information in the public sector. How it is used* (Basingstoke: Palgrave MacMillan).
- W. Van Dooren (2008) 'Nothing New Under the Sun? Change and Continuity in the 20th Century Performance Movement', in W. Van Dooren and S. Van de Walle: *Performance information in the public sector. How it is used* (Basingstoke: Palgrave MacMillan).
- W. Van Dooren, G. Bouckaert and J. Halligan (2015) *Performance management in the public sector*, 2nd edn (London: Routledge).
- S. Van Thiel and F.L. Leeuw (2002) 'The performance paradox in the public sector', *Public Performance and Management Review*, 25, 267–281.
- K. Verhoest (2005) 'Effects of Autonomy, Performance Contracting, and Competition on the Performance of a Public Agency: A Case Study', *Policy Studies Journal*, 33, 235–258.
- A. Wiggins and P. Tymms (2002) 'Dysfunctional effects of league tables: A comparison between English and Scottish primary schools', *Public Money & Management*, 22, 43–48.
- A. Wildavsky (1969) 'Rescuing Policy Analysis from PPBS', *Public Administration Review*, 29, 189–202.
- D.W. Williams (2003) 'Measuring Government in the Early Twentieth Century', *Public Administration Review*, 63, 643–659.
- D.W. Williams (2004) 'Evolution of Performance Measurement Until 1930', *Administration Society*, 36, 131–165.
- K. Wittebrood and M. Junger (2002) 'Trends in violent crime: a comparison between police statistics and victimization surveys', *Social Indicators Research*, 59, 153–173.