Thinking Allowed: Reforming Indicator-Based Accountability to Enhance Innovation

WOUTER VAN DOOREN¹, TOM WILLEMS²

Full reference:

Van Dooren, W., & Willems, T. (2016). Thinking allowed: Reforming indicator-based accountability to enhance innovation. In J. Torfing, & P. Triantafyllou (Eds.), *Enhancing public innovation by transforming public governance* (pp. 273-289). Cambridge University Press.

CONTENTS

1	Introduction	2
2	The Promise and Practice of Indicator-Based Accountability	3
3	Multiple Accountabilities, Indicators and Innovation	5
4	Accountability beyond NPM: A Pro-Innovation Agenda	9
5	Conclusion	13
6	References	14

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 Tom Willems is Postdoctoral Researcher at the research group 'Public Administration & Management', Department Political Science, University of Antwerp, Belgium. Contact: tom.willems@uantwerpen.be

1 INTRODUCTION

Indicator-based accountability was a key tenet of New Public Management (NPM). Public organizations and the people working within those organizations were held to answer for their results rather than for following rules and processes. Unfettered from processual regulation and controls, a blend of creativity and pragmatism was expected to cause better performance and make room for innovation. It turned out differently. NPM-flavoured accountability regimes in many instances stifle innovation rather than foster it. Today, New Public Governance (NPG) creates new challenges for indicator-based accountability systems. NPG raises questions on how to reform indicator-based accountability to make it work in complex multi-actor settings. What is the use of performance indicators when dealing with unruly problems? How to stimulate innovation without calling accountability in itself into question? This chapter proposes some prospective directions based on a critical analysis of indicator-based accountability. We argue that performance indicators should be used for learning and innovation rather than for indicator-based accountability. Accountability can only enhance innovation when it allows for staff and stakeholders to critically think about the results. They have to engage in a performance dialogue on the meaning of indicators that also accounts for complexity. Indicator-based accountability, however, typically does not allow for such thinking. The target is the truth. In our view, reforming accountability therefore implies that results-based accountability needs to be saved from its indicators.

Accountability is foundational to all variants of governance: the Classical Public Administration (CPA), NPM and NPG. In each governance paradigm, however, accountability takes on a particular interpretation. While CPA accountability is based on rule-following in a hierarchy, NPM accountability focuses on results in a (quasi-)market. In NPM, results-based accountability is typically narrowed down to indicator-based accountability. One of the catchphrases of NPM is that only what is measured matters. NPM hence assumes that we cannot talk about results in a qualitative, judgemental way. In this chapter, we look at indicator-based accountability as a subset of results-based accountability. The accountability model in NPG is more diffuse than the CPA and NPM approaches (Willems and Van Dooren 2011). There is some rule-following of general principles rather than detailed regulations. There is a result orientation but at a cross-cutting, more general level. NPG accountability also puts a stronger emphasis on accountability mechanisms within professional groups. In what follows, we will first critically address NPM's indicator-based accountability. We then will argue that indicator-based accountability is entangled in a diffuse web of multiple accountabilities, which bars the prospect for innovation. Finally, we discuss how indicators can foster innovation. We propose not to use indicators for accountability but to enrich professional learning and dialogue. Performance measurement should build on intrinsic motivation rather than be an external control. This view is more consistent with the NPG paradigm.

2 THE PROMISE AND PRACTICE OF INDICATOR-BASED ACCOUNTABILITY

The NPM diagnosis of the established bureaucracies under the CPA model was straightforward. The excessive density of rules and regulations obstructed all creativity and pragmatism at the front line of public services. Read the examples of Osborne and Gaebler's (1993) acclaimed text on NPM. The parks department in Visalia, California, could provide a deposit of \$60,000 overnight to buy a \$400,000 castoff swimming pool because budget rules allowed for managerial freedom. A defaulting community school in East Harlem became the best in class after the district did away with assignment by zone. Free school choice created a quasi-market and led to competitive pressures to improve. In the US Department of Defense, a deputy assistant secretary for installations cut the rules for military base construction from 400 to 4 pages and those for housing from 800 to 40. Base commanders could make the decisions autonomously now. As an experiment, one military base was even completely relieved from regulation. The gains could be reinvested in the base. As a result, thousands of regulations have been put into question, costs have been cut and performance increased 3 per cent. Efficiency gains were believed to be 10 per cent or \$3 billion. The driver in all cases was front-line innovation.

The remedies of the NPM advocates promised the best of both worlds: fewer rules, more innovation, matched with better accountability. More flexibility would lead to creativity, innovation and better performance. Account holders would retreat for months, years or any other time-to-target. Relieved of continuous account giving on processes and the use of resources, public managers would be more prepared to take the risks needed for innovation. Yet, this did not imply that politics had to clear the way. On the contrary, politicians would be able to focus on what 'really matters', and what really matters are results, not rules.

In reality, the bifurcation between NPM and CPA was clearly not as sharp as presented. Well before NPM, the implementation literature had amply demonstrated the importance of front-line discretion (Hill and Hupe 2002). Similarly, indicator-based steering was not an invention of the managerial movement of the 1990s. The application of indicator-based management, however, has never been as intensive, i.e. integrated in the incentive structures of daily management, and extensive, i.e. applied in virtually all corners of the public sector (Van Dooren 2008).

The practice of indicator-based accountability proved to be challenging. Radin (2000) succinctly summarized her critique of NPM arguing that it is fitting square pegs in round holes. Rather than a failure of implementation, indicator-based accountability is flawed in its design: the main default being its naïve view on information. NPM audaciously assumed the forthright availability of performance information (Bouckaert and Peters 2002). Measurement of performance is a critical feature of results-based accountability

schemes. Yet, as Osborne and Radnor argue in Chapter 3, robust measurement is only feasible in production environments with tangible routine tasks (Noordegraaf and Abma 2003). Precisely those routine tasks have become increasingly rare in public services for at least three reasons. First, wicked policy issues lead to more complex service-delivery arrangements and less routine. This is most evident in the services dealing with multi-problems. Think of the drug addict who is also homeless, unemployed and poorly educated. Effective service delivery for multi-problems relies on networking and counselling rather than the mere production of methadone programmes, shelter and training courses. Second, services are increasingly embedded in professional disciplines and subject to professional expertise that objects routine work. Few professionals would or could describe their work in production terms (Jos and Tompkins 2004). Third, many (but not all) constituencies have been outsourcing the remainder of routine tasks in government to the private sector, which further reduce the prospect of sensible performance measurement in the public realm (Radin 2006).

Also the assumption that results-based accountability schemes would substitute existing rules and regulations proved to be naïve and partly flawed. Poulsen (2009) for instance describes the co-existence of competing public administration traditions, with often contradictory interpretations of accountability, as different 'archaeological layers'. Rather than substitution, regulation is a historical sedimentation process in which new mechanisms add onto older ones (see also Lægreid and Verhoest 2010; Schillemans 2010; Van Dooren, Voets and Winters 2015). Indicator-based accountability proved to be just another layer to the sediment of regulation.

The practice of the next decennium proved the critics right. People in public services resorted to gaming the performance metrics (see Pollitt 2013 for an overview). Gaming of indicators in high-stakes accountability schemes is rampant; police officers reclassify crimes, researchers set up citation circles, teachers teach for the test and hospitals reduce waiting lists by operating a waiting list for the waiting list. The options for creative accounting at the front line seem endless (Bohannon 2013; Van Dooren, Bouckaert and Halligan 2015). If the gamer were only a statistics juggler, not too much harm may be done. Yet, this is often not the case. Indicator-based accountability can affect the core processes and outcomes of public services. Several studies show that the impact on front line work is complex, subtle, but definitely real. A study by Soss, Fording and Schram (2011: 225) in a Florida welfare programme falling under an indicator-based accountability scheme found that 'the discretion possessed by case managers is broad, in the sense that they are authorized to make a wide variety of decisions affecting the client, and it is ineradicable, in the sense that they almost always know some way to push a decision in a preferred direction'. But, they argue, the broad discretion does not run very deep. Time pressure of the performance regime prevents front line workers from pursuing the treatments they would have advised otherwise. Brodtkin's (2011) study of welfare reform introduces the concept of street-level calculus. Street-level practitioners do more than simply respond to performance incentives; they also adjust to them as they

manage the imbalance between the demands of their jobs, the indicators and resource constraints.

3 MULTIPLE ACCOUNTABILITIES, INDICATORS AND INNOVATION

But why is indicator-based accountability leading to gaming and rigidity rather than to service improvement and innovation? In order to answer this question, we first need to refine our understanding of accountability. We argue that accountability takes place in five accountability forums that host the main processes of contemporary accountability (Table 1). We take the metaphor of a forum quite literally – an approach which is slightly different from the functional approach used in Chapter 15 by Tom Christensen and Per Læg Reid (see also Schillemans and Bovens 2011). A forum is a marketplace where actors are held to account by account holders, following the rules of that particular marketplace. Account givers and account holders typically belong to a specific forum: administrative agencies and cabinet ministers for instance to the administrative forum or media players to the public forum. However, willingly or not, account givers are regularly drawn to other forums: for instance when an agency has to answer to court for violating legislation or when the media challenges a minister. Since public–private collaboration is increasingly strong, we include the market as a relevant forum where public account givers have to answer for their actions.

Table 1: Forums and processes of accountability

Accountability forums	Processes of accountability
Political	Elections, parliamentary scrutiny, political debate within political parties, ministerial responsibility,...
Administrative	Hierarchical command, diverse government auditors, regulatory bodies, ombudsmen, other types of government monitoring agencies, performance contracts, budgets,...
Judicial	Judicial courts, administrative tribunals, mediation,...
Public	Mass media, social media, civic action,...
Market	Shareholders' reports, consumer choice, rating agencies,...

(Source: Willems and Van Dooren 2012)

Indicator-based accountability is developed in the administrative forum, where a principal 'lets' and 'makes' managers manage through the use of performance contracts with targets. During the time-to-target, the principal unloads most accountability demands. The period in between contract evaluations allows at least in theory for risk-taking and innovation. Without the permanent immanence of accountability, accountees have the time to correct

for experiments and innovations that went wrong. Yet, the impact of indicators transcends the principal-agent setting of the administrative forum. Indicators play out in other forums in a very unpredictable way. Politicians, journalists and investors for instance do not follow the orderly cycles of performance contracting. Rather, they grab the indicators they need, when they need them. The oxygen for risk-taking and innovation that performance contracts are supposed to offer is drained off again, because indicators can be used in other forums than the administrative forum at any moment.

The combined action of multiple accountabilities can be seen as a case-specific field of forces that creates incentives for rule following, improvement, gaming or any other organizational behaviour, including the behaviour that stimulates innovation. These multiple accountabilities define the interests at stake in performance indicators, and in second order also define the freedom of action that is needed for innovation. In many, but not all contexts, the use of performance indicators for multiple accountabilities leads to indicator rigidity rather than an environment supportive to innovation.

Indicator-based accountability is primarily developed within the *administrative forum*. Executive politicians hold public managers to account for results. The whole endeavour rests on performance indicators that are typically found in institutional instruments such as annual reports, management contracts, scorecards and performance budgets. Performance management in the public sector is moreover becoming a specialization in itself, with increasingly more staff devoted to measurement, evaluation and auditing (Van Dooren 2008). Interestingly, there is usually not much political opposition against performance-based management reform. In many countries, NPM reforms have been approved quasi-unanimously in parliament (Pollitt and Bouckaert 2004). The issue is depoliticized, it seems.

Accountability processes within the administrative forum can perfectly explain indicator rigidity that stands in the way of innovation. A rigid focus on indicators occurs when performance targets are consequential for individual careers. Performance pay schemes have been demonstrated to have a strong impact on gaming behaviour (Perry, Engbers and Jun 2009). The detrimental effect of the bonus culture in the financial sector is a case in point for the displacement effect of performance incentives. Rigidity also occurs when performance targets have an impact on budgets of organizations (Hood 2006). When money is tied to the results, the temptation to juggle the indicators is high.

Yet, we cannot suffice by looking only at the administrative forum. In many instances, performance indicator schemes are noncommittal at first sight. Performance scorecards for instance only make information available without targets, punishments or rewards. Nonetheless, rigidity regularly occurs in such apparently low-pressured accountability schemes (Wiggins and Tymms 2002; Hood 2006). Moreover, many seemingly high-pressured performance schemes such as performance pay and budgeting turn out to be relatively low-pressured because sanctions often fail to materialize (Schick

2003). Yet, even when high pressure becomes low pressure, gaming is looming because performance indicators unfold in other forums. When indicators are used in other accountability forums, they also follow the rules of these forums.

Performance indicators play a somewhat dubious role in the political forum. There is ample evidence that Members of Parliament or local councils use performance information rather reservedly, if at all (Johnson and Talbot 2008; Van Dooren, Bouckaert and Halligan 2015). On some occasions, however, there are outbursts of political debate on performance metrics: on punctuality of trains for instance, or educational attainment and the incidence of hospital bacteria. Suddenly, performance indicators are politically significant. Yet, rather than following the orderly march of administrative reporting, performance indicators are frequently used as ammunition in political guerrilla warfare. In the garbage can of policy making, political players can use performance indicators to make problems politically relevant at any time (Cohen, March and Olsen 1972; Kingdon 1984).

Indicators also play out in the public forum. News value generally increases when stories are framed negatively (Galtung and Ruge 1965). The performance literature refers to this phenomenon as negativity bias (James and John 2007). The reputational impact of public exposure hence can be substantial, and public organizations devote considerable efforts to avoid public blame (Hood 2010). Arguably, blame management has shifted gears with the surge of social media that cause public blame to go viral. Media exposure of performance targets has become increasingly unpredictable and the threat of bad press is always in the air. Performance indicators can certainly fuel the blame game (Hood 2010).

In addition to media and politics, there is also accountability in markets or quasi-markets. Quasi-markets that are driven by league tables of school performance are a good example. The performance information is supposed to inform parents on school choice. The policy theory behind these reforms has been typified as the choice revolution (Blomqvist 2004). The provision of an exit option is supposed to put pressure on public services. The choice discourse was strong in Sweden, the United Kingdom and the Netherlands. The European Union used the concept of choice to argue for liberalization of public services (Jilke and Van de Walle 2012). Market pressure can be quite substantial and a good deal of gaming occurs on league table indicators. The teaching-for-the-test tactic is probably one of the best-known instances of gaming (Hood 2006). Gaming should not come as a surprise. Public services are concerned that league tables may lure them in a vicious spiral. Fewer consumers may opt for a particular public service, which in turn may lead to demotivation and scale inefficiencies. For human services such as education, the vicious circle also affects the intake of the public service. A high-ranking school can be selective at the gate. At the same time, the low-ranking schools are left with the deprived pupils.

Accountability processes hosted by the legal forum are seldom a direct cause of performance indicator rigidity. Failures to deliver upon performance targets can rarely be taken to court. But legal obligations have an impact in a different way. They can explain why regularly no use is made of the discretion, if any, that a focus on results would create. The use of discretion to reach targets may also lead to legal liabilities. The use of discretion holds a risk of being held accountable in court for failure. Sticking to the rules and routines is generally the safer option (Bardach and Kagan 1982). Clearly, the risks of risk-taking are found in all governance arrangements that rely on front-line discretion.

The different demands for accountability point organizations towards an inflexible, rigid dealing with performance indicators. Whereas the hope within the administrative forum is that accountability for results creates room for innovation, other forums of accountability lead organizations to treat indicators as cast in concrete. Whereas the administrative principal-agent logic proposes that managers are free to manage as long as they measure results, the message heard from the other accountability forums is that everything they measure can be used against them, anytime.

Why does indicator-based accountability often stand in the way of innovation? The answer is essentially the same as why old-fashioned bureaucracy can obstruct innovation. Indicator-based accountability discourages risk-taking in several ways. First, the reputational damage that could result from bad media coverage leads public organizations to a more narrow focus on indicators. Second, in quasi-market contexts, pressure further increases. Teachers who mainly teach for a test are a case in point. Bad results may lead to lower enrolment and hence in the long run threaten organizational survival. Third, political blame games may lead to bureaucracy bashing when targets are not met, but also to political recuperation when targets are reached. Fourth, organizations often fall back on indicators of processes when output or outcome measures are not available. In this case, experimentation with processes becomes difficult. Fifth, when innovation leads to higher performance, it may trigger a ratchet effect (Hood 2006). Targets are typically set at a higher level when performance increases. This creates an incentive to ensure not to overshoot the targets because surpassing them would lead to higher targets the next year (Van Thiel and Leeuw 2002). After performance targets are introduced in a scheme of accountability, results tend to cluster around the target. Low performers make efforts to reach the target, but high performers tend to relax their effort. Finally innovation often results in an S-shaped performance improvement curve. Performance peaks immediately after introducing an innovation but declines when the innovation momentum fades away. Due to ratchet effects, innovators fear the immediate dip in performance after innovation since it is penalized in the target regime.

4 ACCOUNTABILITY BEYOND NPM: A PRO-INNOVATION AGENDA

Performance information will only lead to learning and innovation when the pressure emanating from indicator-based accountability is driven under the gaming threshold. As we argued above, we have reasons to believe that this threshold is rather low. The discussion of accountability forums suggests that performance indicators can be taken to the political and public forum any time. It has also been suggested that political and public accountability follow a garbage can model that makes accountability for performance unpredictable. Real, biting challenges of accountability rarely coincide with formal cycles of management reporting. The unpredictability creates uncertainty and risk-avoidance, which in turn discourages experimentation.

So how to reform indicator-based accountability to enhance innovation? The first and foremost reform of accountability mechanisms to enhance innovation would be to take performance *indicators* out of the formal processes of accountability, such as budgeting, contracting and performance appraisal. Indicator-based accountability should move away from formal schemes that usually lead to compliance at best. What are needed are performance indicators that trigger learning. Performance information can trigger the disruptive questions that lead to innovation. The high scores for Finland on the OECD education rankings, for instance, should challenge educational policies. They should not lead to naïvely copying the Finnish education system or to a political blame game in low-performing countries (Meyer and Benavot 2013). The education indicators should trigger policy learning. This requires a deeper understanding of what is going on. Is it really the structure of the education system? Or is the schooling of the teachers better in Finland? Are class sizes maybe smaller or is the student population less diverse? Policy learning should also critically question the value of the indicators themselves. What do comparative tests actually measure for instance? Is it mainly cognitive test results? And if so, what about values such as creativity or citizenship? The same goes for indicators within organizations. Lotte Bøgh Andersen (Chapter 12) shows that command and incentive systems can work only when transformative leadership enables dialogue on the indicators. Indicator-based accountability schemes, however, often do not make room for asking challenging questions. Explaining the past is more important than preparing for the future (Carter, Klein and Day 1992).

Cutting back on indicator-based accountability offers an exiting prospect for performance management. Contemporary performance management is often reduced to performance *indicator* management. This has led to many adverse effects that fundamentally challenge the utility of performance management. Yet, performance management is more than performance measurement. It implies that managers look at results and outcomes rather than inputs and processes. Performance managers in theory attach great importance to the decentralization of responsibilities; as long as results are

obtained managers are free to manage, and staff are free to organize their work. Similarly, results-based accountability is more than indicator-based accountability. Many results, and particularly innovative results, are hard to quantify beforehand but nonetheless very real. Contrarily, many quantified results lack meaning in practice. Genuine performance management should build on meaningful conceptions of performance.

When management no longer prescribes how results need to be obtained, high trust is placed in front-line workers to do their job. A focus on results offers an opportunity to engage with staff to make meaning of performance. Arguably, people are quite willing to account for results, but only when they can do so in a meaningful way. Accountability for results can only work when real, meaningful results and impacts are discussed. This is not the case in many indicator-based accountability schemes, as the literature on gaming extensively demonstrates. Such schemes subvert the trust in front-line workers and professionals that is needed to make performance management work (Soss, Fording and Schram 2011). Indicators can also enable what Noordegraaf (2008) calls discursive craftsmanship. Indicators play a role in crafting policy stories that are considered to be meaningful. The metaphorical power of numbers to build meaningful policy stories has also been documented by Stone (1997). Performance management is mainly about sense making. Many performance indicators, however, are not making sense.

We do not suggest abandoning performance indicators altogether. In order to save performance management from its indicators, new connections need to be made. Performance management systems that tightly couple measurement to evaluative judgements need to be replaced by loosely coupled systems that allow for deliberation, learning and ultimately innovation (Van Dooren 2006; Van Dooren, Bouckaert and Halligan 2015). Performance management scholars have already studied this new approach to performance indicators. Moynihan (2008) proposes to embed performance indicators into a performance dialogue that allows for interpretation and learning. A performance dialogue, according to Moynihan, is a routine event that takes all relevant stakeholders on board. The main purpose is to make meaning from performance indicators by using both quantitative and experiential data. The discussion should be non-confrontational and be based on equality among participants. Performance dialogues should focus on goals, but could also serve as a breeding ground for innovation. The dialogue on performance should make sense of the mostly ambiguous performance information. Kroll (2013) adds to this notion of routine dialogues the idea of non-routine feedback in performance management. Much of the value of performance indicators does not originate from the indicator-based accountability schemes. What is more important are these spontaneous mechanisms of feedback that lead to ad hoc learning.

Cutting down on indicator-based accountability would by no means be the end of results-based accountability either. Indicator-based accountability is a limited approach to results-based accountability. Forums of accountability can perfectly thrive and be results-oriented without catalogues of perfor-

mance indicators (Bovens 1998; Willems and Dooren 2011). Accountability mechanisms such as legal appeal and media exposure, hierarchy and market exit are still in place. Politicians can still question the performance of the education system when the PISA study comes out. Media can and will still critique the railway company when punctuality of the trains is lacking. Performance audits are still published and interest groups are still issuing standpoints. These external pressures can be beneficial for innovation. Yet, as we argued above, large and high-pressured performance indicator schemes are generally at odds with innovation. When schools are held accountable for test scores, teachers teach the test. Similarly, there are many ways in which railway companies can improve punctuality without improving the quality of service. Public organizations may even use indicators to divert questions of accountability, rather than to answer them. In this way, indicators shield off pressures for innovation rather than take innovation to heart.

NPG aspires to respond to complex societal contexts. Faced with complexity, results-based accountability needs thick descriptions of performance to allow for performance dialogue and meaningful innovation. The main critique on indicator-based accountability has targeted the thin evidence base of performance indicators. As the gaming literature has demonstrated abundantly, performance indicators are often not valid because they only measure performance very partially. They are also not reliable because strategic behaviour skews measured results. Rather than publicizing a laundry list of performance indicators, public services could bring a documented evaluation of performance and goal attainment to the table. The development of thick descriptions comes with a warning, however. Lewis and Triantafyllou (2014) argue that thick description is more likely to supplement than to replace existing performance indicator regimes of accountability; that it is likely to demand more not less data; that it requires extensive participation and dialogue; and that it may reinforce the logic of constant organizational change in the name of improvement. They warn that such 'learning systems' may simply increase administrative overload rather than reduce it. The layering of oversight mechanisms may indeed suffocate innovation capacity and should be avoided.

Innovation typically builds on smart information gathering, with a good combination of both thin and thick descriptions. Thin performance indicators may be used to scratch the surface and to trigger debate. They may also direct our attention to issues that are worthwhile of thick description. Thick descriptions are better suited to finding explanations for differences in performance and for designing future strategies. Thick descriptions may also raise critical voices on thin performance indicator schemes. A good performance dialogue should reveal the instances of gaming that may have affected the performance data. In fact, most developed countries have a longstanding tradition of policy evaluation. The evidence base for innovation will be strengthened when evaluation studies and performance indicators are both put to use. Smart combinations of thick and thin evidence on performance should also be useful for scrutiny on different forums of accountability.

A step further would be to hold organizations and programmes accountable for innovation rather than for results. The innovation literature offers some clues on the responsibilities for which organizations could be held accountable: creating an open environment that fosters the intake and circulation of ideas, providing for a safe context for change, attention for diversity in the organization, scanning of the environment, prototyping and piloting (Albury 2005; Hartley 2005). These dimensions describe the capacity for innovation rather than the occurrence of innovation or the impact of innovation on society. More capacity assumes a higher *probability* of relevant innovation. The occurrence of innovation itself is too unpredictable and the impact of innovation is too diffuse to be a basis for accountability (Osborne and Brown 2013). Clearly, it will be hard to quantitatively measure capacity for innovation. We should not fall victim to the same mistakes of traditional indicator-based management. Are we able to quantitatively measure an inherently complex and context-dependent concept such as innovative capacity? Thick descriptions of innovative capacity might be more meaningful. Mazzucato (2013), for instance, critiques the number of patent registrations as an indicator of private sector innovativeness. She demonstrates that while private companies register most patents, the most innovative work comes out of the public sector, with government labs and government-backed universities investing in the research responsible for producing the most radical new drugs. Clearly, a thick description that assesses the quality of the drugs is needed to really understand innovation in the pharmaceutical sector.

Third, chance events seem a particularly relevant subject for both accountability and innovation (see also Chapter 13 by Montgomery Van Wart). Political crises, natural disasters and policy failures put pressure on the resilience of the political and administrative system. Chance events also activate processes of accountability on all forums (Schneider 2005; Rixen 2013). A crisis is a lightning switch for all accountability forums. Recent terrorist attacks in Paris and Copenhagen, for instance, led to public scrutiny of the role of government in politics, media and the courts. After a foiled terrorist attack in Belgium, a discussion ensued on quasi-market 'voting with the feet' by Jewish communities from allegedly inadequate protection by police forces. A key issue is whether accountability pressure leads to innovation (Osborne and Brown 2013). In any case, processes of accountability do have the potential to translate disruptive events into discontinuous change. Although not everyone agrees that it has been an innovation for the better, the 9/11 attacks without any doubt led to such a major overhaul of public governance (Roberts 2011).

Fourth, a politicization of accountability mechanisms may be fruitful. Governance mechanisms in general and performance indicator regimes in particular are only to a limited extent scrutinized in political and public forums. The news values and political salience of policy content are generally higher than the governance arrangements for these policies (Roberts 2011; Willems and Van Dooren 2016). Yet, liberal democracies should be capable of having a political debate on the advantages and disadvantages of accountability arrangements and decide on how much and what kind of account-

ability is required. Until now, mainly scandals have sparked off a politicized debate on indicator-based accountability. Patient neglect at Stafford Hospital (United Kingdom) for instance has been attributed to indicator rigidity and accountability (Mid Staffordshire NHS Foundation Trust Public Inquiry 2013). In the United States, a public debate on performance incentives unfolded after Veteran Affairs hospitals were scrutinized for disregarding patient follow-up (Moynihan 2014). A more sustained and systematic review of accountability mechanisms may be useful. Such a review should go beyond scandals and investigate the many functions of accountability; including innovation.

5 CONCLUSION

The overall argument of this chapter has been that indicator-based accountability, as proposed by the NPM paradigm, often drives out innovation. The promise of freedom to manage and to innovate does not materialize due to a rigid focus on performance indicators. NPM thus reduces results-based accountability to indicator-based accountability. Moreover, NPM only acknowledges the administrative forum of accountability, where (political) principals set targets for bureaucrats and agencies. In reality, organizations face multiple accountabilities. Performance indicators thus play out in different forums of accountability such as politics, media and markets. Indicators are used in an unpredictable way. Overall, it seems that indicator-based accountability following the NPM logic insufficiently acknowledges the complexity of contemporary governance. The simplicity of the model – set target, measure and sanction – is often a barrier to innovatively solving complex policy issues.

In contrast to NPM, NPG takes complexity as its problem diagnosis. Innovation can only materialize in governance systems that cope with (and thrive in) complex settings. A reform of indicator-based accountability can only enhance innovation if it is able to deal with complexity too. We argue that performance indicators can help to understand complexity when used for a performance dialogue outside of an accountability context. This is a learning perspective that has a better probability of understanding complexity and getting to innovation. Furthermore, indicator-based accountability may benefit from a good combination of thick and thin description. Another suggestion is to demand accountability for innovation rather than results. Since innovation typically is unpredictable, the focus should be on innovation capacity rather than actual innovations. Finally, accountability and innovation should be politicized. Thriving accountability forums in a liberal democracy should not only discuss the contents of policies but also the nature of accountability and the value of innovation.

6 REFERENCES

- Albury, D. 2005. 'Fostering innovation in public services', *Public Money & Management* 25(1): 51–6.
<http://doi.org/10.1111/j.1467-9302.2005.00450.x>.
- Bardach, E. and Kagan, R. A. 1982. *Going by the Book the Problem of Regulatory Unreasonableness*. Philadelphia: Transaction Publishers.
- Blomqvist, P. 2004. 'The choice revolution: Privatization of Swedish welfare services in the 1990s', *Social Policy & Administration* 38(2): 139–55.
<http://doi.org/10.1111/j.1467-9515.2004.00382.x>.
- Bohannon, J. 2013. 'Who's afraid of peer review?' *Science* 342(6154): 60–5. <http://doi.org/10.1126/science.342.6154.60>.
- Bouckaert, G. and Peters, B. G. 2002. 'Performance measurement and management: The Achilles' heel in administrative modernization', *Public Performance & Management Review* 25(4): 359–62.
- Bovens, M. 1998. *The Quest for Responsibility: Accountability and Citizenship in Complex Organisations*. Cambridge, MA: Cambridge University Press.
- Brodtkin, E. Z. 2011. 'Policy work: Street-level organizations under new managerialism', *Journal of Public Administration Research and Theory* 21(S2): i253–77. <http://doi.org/10.1093/jopart/muq093>.
- Carter, N., Klein, R. and Day, P. 1992. *How Organizations Measure Success. The Use of Performance Indicators in Government*. London: Routledge.
- Cohen, M. D., March, J. G. and Olsen, J. P. 1972. 'A garbage can model of organizational choice', *Administrative Science Quarterly* 17(1):1–25.
<http://doi.org/10.2307/2392088>.
- Galtung, J. and Ruge, M. H. 1965. 'The structure of foreign news: The presentation of the Congo, Cuba and Cyprus crises in four Norwegian newspapers', *Journal of Peace Research* 2(1):64–90.
<http://doi.org/10.1177/002234336500200104>.
- Hartley, J. 2005. 'Innovation in governance and public services: Past and present', *Public Money and Management* 25(1): 27–34.
<http://doi.org/10.1111/j.1467-9302.2005.00447.x>.
- Hill, M. and Hupe, P. 2002. *Implementing Public Policy: Governance in Theory and in Practice*. London: Sage.
- Hood, C. 2006. 'Gaming in targetworld: The targets approach to managing British public services', *Public Administration Review* 66(4): 515–21.
- Hood, C. 2010. *The Blame Game: Spin, Bureaucracy, and Self-Preservation in Government*. Princeton: Princeton University Press.

- James, O. and John, P. 2007. 'Public management at the ballot box: Performance information and electoral support for incumbent English local governments', *Journal of Public Administration Research and Theory* 17(4): 567–80.
- Jilke, S. and Van de Walle, S. 2012. 'Two track public services? Citizens' voice behaviour towards liberalized services in the EU15', *Public Management Review* 15(4): 465–76.
<http://doi.org/10.1080/14719037.2012.664015>.
- Johnson, C. and Talbot, C. 2008. 'UK parliamentary scrutiny of public services agreements: A challenge too far?', in S. Van de Walle and W. Van Dooren (eds.), *Performance Information in the Public Sector: How it is Used*. Basingstoke: Palgrave MacMillan, pp. 140–57.
- Jos, P. H. and Tompkins, M. E. 2004. 'The accountability paradox in an age of reinvention: The perennial problem of preserving character and judgment', *Administration & Society* 36(3): 255–81.
- Kingdon, J. 1984. *Agenda, Alternatives and Public Policies*. Boston: Little, Brown & Company.
- Kroll, A. 2013. 'The other type of performance information: Nonroutine feedback, its relevance and use', *Public Administration Review* 73(2): 265–76. <http://doi.org/10.1111/j.1540-6210.2012.02648.x>.
- Lægreid, P. and Verhoest, K. 2010. 'Reforming public sector organizations', in *Governance of Public Sector Organizations: Proliferation, Autonomy and Performance*. Basingstoke: Palgrave Macmillan, pp. 1p.
- Lewis, J.M. and Triantafillou, P. 2014. 'From performance measurement to learning: a new source of government overload?' *International Review of Administrative Sciences* 78(4): 597–614.
- Mazzucato, M. 2013. *The Entrepreneurial State: Debunking Public vs. Private Sector Myths*. New York: Anthem Press.
- Meyer, H.-D., and Benavot, A. 2013. *PISA, Power, and Policy: The Emergence of Global Educational Governance*. Oxford: Symposium Books.
- Mid Staffordshire NHS Foundation Trust Public Inquiry. 2013. *Report of the Mid Staffordshire NHS Foundation Trust Public Inquiry: Executive Summary*. The Stationery Office.
- Moynihan, D. 2014. 'The problem at the VA: "Performance perversity"', *Los Angeles Times* (June 1). www.latimes.com/opinion/op-ed/la-oe-moynihan-va-scandal-performance-perversity-20140602-story.html.
- Moynihan, D. P. 2008. *The Dynamics of Performance Management: Constructing Information and Reform*. Washington, DC: Georgetown University Press.
- Noordegraaf, M. 2008. 'Meanings of measurement', *Public Management Review* 10(2): 221–39. <http://doi.org/10.1080/14719030801928672>.

- Noordegraaf, M. and Abma, T. 2003. 'Management by measurement? Public management practices amidst ambiguity', *Public Administration* 81(4): 853–71.
- Osborne, D. and Gaebler, T. 1993. *Reinventing Government: How the Entrepreneurial Spirit is Transforming the Public Sector*. New York: Plume Press.
- Osborne, S. P. and Brown, K. 2013. *Handbook of Innovation in Public Services*. Cheltenham: Edward Elgar.
- Perry, J. L., Engbers, T. A. and Jun, S. Y. 2009. 'Back to the future? performance-related pay, empirical research, and the perils of persistence', *Public Administration Review* 69(1): 39–51.
- Pollitt, C. 2013. 'The logics of performance management', *Evaluation* 19(4): 346–63.
- Pollitt, C. and Bouckaert, G. 2004. *Public Management Reform: A Comparative Analysis*. Oxford: Oxford University Press.
- Poulsen, B. 2009. 'Competing traditions of governance and dilemmas of administrative accountability: The case of Denmark', *Public Administration* 87(1): 117–31.
- Radin, B. A. 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(1): 111–35.
- Radin, B. A. 2006. *Challenging the Performance Movement: Accountability, Complexity, and Democratic Values*. Washington, DC: Georgetown University Press.
- Rixen, T. 2013. 'Why reregulation after the crisis is feeble: Shadow banking, offshore financial centers, and jurisdictional competition', *Regulation & Governance* 7(4): 435–59. <http://doi.org/10.1111/rego.12024>.
- Roberts, A. 2011. *The Logic of Discipline: Global Capitalism and the Architecture of Government*. Oxford: Oxford University Press.
- Schick, A. 2003. 'The performing state: Reflection on an idea whose time has come but whose implementation has not', *OECD Journal on Budgeting* 3(2): 71–103.
- Schillemans, T. and Bovens, M. 2011. 'The challenge of multiple accountability: Does redundancy lead to overload?', in M. Dubnick and G. Frederickson (eds.), *Accountable Governance: Problems and Promises*. New York: ME Sharpe, pp. 3–21.
- Schneider, S. K. 2005. 'Administrative breakdowns in the governmental response to Hurricane Katrina', *Public Administration Review* 65(5): 515–6.

- Soss, J., Fording, R. and Schram, S. F. 2011. 'The organization of discipline: From performance management to perversity and punishment', *Journal of Public Administration Research and Theory* 21(s2): i203–32. <http://doi.org/10.1093/jopart/muq095>.
- Stone, D. 1997. *Policy Paradox: The Art of Political Decision Making*. New York: W. W. Norton and Company.
- Van Dooren, W. 2006. *Performance Measurement in the Flemish Public Sector: A Supply and Demand Approach*. Thesis, faculty of social sciences. Leuven, Belgium.
- Van Dooren, W. 2008. 'Nothing new under the sun? change and continuity in the twentieth century performance movement', in S. Van de Walle and W. Van Dooren (eds.), *Performance Information in the Public Sector: How It Is Used*. Basingstoke: Palgrave MacMillan, pp. 15–45.
- Van Dooren, W., Bouckaert, G. and Halligan, J. 2015. *Performance Management in the Public Sector*, 2nd edition. London: Routledge.
- Van Dooren, W., Voets, J. and Winters, S. 2015. 'Autonomy and reregulation: Explaining dynamics in the Flemish social housing sector', *Public Administration* 93(4), 1068–1083.
- Van Thiel, S. and Leeuw, F. L. 2002. 'The performance paradox in the public sector', *Public Performance & Management Review* 25(3): 267–81.
- Wiggins, A. and Tymms, P. 2002. 'Dysfunctional effects of league tables: A comparison between English and Scottish primary schools', *Public Money and Management* 22(1): 43–8.
- Willems, T. and Van Dooren, W. 2011. 'Lost in diffusion? How collaborative arrangements lead to an accountability paradox', *International Review of Administrative Sciences* 77(3):505–30. <http://doi.org/10.1177/0020852311408648>.
- Willems, T. and Van Dooren, W. 2012. 'Coming to terms with accountability', *Public Management Review* 14(7): 1011–36. <http://dx.doi.org/10.1080/14719037.2012.662446>.
- Willems, T. and VanDOoren, W. 2016. '(De)Politicization dynamics in public-private partnerships (PPPs): Lessons from a comparison between UK and Flemish PPP policy', *Public Management Review* 18(2): 1–22. <http://doi.org/10.1080/14719037.2014.969759>.