



Sense and complexity: Initiatives in responsive performance audits

Evaluation
17(4) 351–363
© The Author(s) 2011
Reprints and permission: sagepub.
co.uk/journalsPermissions.nav
DOI: 10.1177/1356389011423551
evi.sagepub.com



Peter van der Knaap

Netherlands Court of Audit, The Netherlands

Abstract

Many policy programmes yield disappointing results. Like policy evaluation, performance audit aims to help policy makers by uncovering barriers to success and suggesting improvements. Traditionally, performance audit relies heavily on policy objectives, norms, and criteria while focusing on the responsibility of a specific government entity (e.g. a ministry or an agency). More recently, however, performance auditors have recognized the need to be more open and responsive to the needs and preferences of target groups and stakeholders, and changes in context and knowledge – elements that have been central to responsive and realist evaluation for some years. This article describes new initiatives to bring more responsiveness into performance audits. Starting from the new international performance audit guidelines of INTOSAI, it discusses how this can work in practice by looking at recent initiatives of the Netherlands Court of Audit. The focus is on a new tool for performance audit, under the rather provocative label of ‘reality checks’: compact audits of the connection between policy programmes and stakeholders’ needs and preferences. By including more responsive research questions and methods and looking critically to existing monitoring and evaluation practices in policy programmes, performance audit has an important role to play in balancing the benefits and risks of policy objectives and performance indicators.

Keywords

indicators, learning; objectives, performance audit, performance management, responsive evaluation

Abandon the urge to simplify everything, to look for formulas and easy answers, and to begin to think multidimensionally, to glory in the mystery and paradoxes of life, not to be dismayed by the multitude of causes and consequences that are inherent in each experience; to appreciate the fact that life is complex.
(M. Scott Peck)

Corresponding author:

Peter van der Knaap, Netherlands Court of Audit, Directorate for Performance Audit, Lange Voorhout 8, NL-2514 ED, The Hague, The Netherlands.

Email: peter.vanderknaap@rekenkamer.nl

Introduction

The editor of this journal has suggested that audit, and particularly performance audit, converges with evaluation in many ways (Stern, 2011). Still, there seems to be a love–hate relationship between the two. When I asked participants of a 2008 EES Conference session to either cheer or boo in response of words I had projected out on a big screen, you can imagine the response to ‘audit’. Words like ‘bean counting’ and ‘nitpicking’ often come into play when performance audit is being discussed by evaluators (Reichborn-Kjennerud and Johnsen, 2011). Many evaluators, however, work in audit institutions and these same audit institutions play an important role in evaluation societies and practices around the world. In addition, the strong institutional position of national audit institutions allows for new and even unexpected initiatives that may be relevant to more than just the ‘audit society’ (Power, 1997). This may be especially true as performance audit, while keeping its edge, tries to find ways to mitigate the negative consequences of their compliance-oriented craft.

This article is about initiatives that fall into this category: those to take a more responsive approach in performance audit. It is structured as follows. Section 2 describes the background and theoretical framework of performance audit. It makes a case for the ‘sense’ of using norms and criteria in both policy and audit. Section 3 points to the risks: over-simplification, tunnel vision, resistance to change, and an increasing lack of relevance. It describes the call for responsiveness in evaluation and audit literatures. In Sections 4 and 5 recent initiatives in The Netherlands to make performance audit more responsive are described, with a focus on what is labelled ‘reality checks’: short, bottom-up investigations into the relevance and actual appreciation by stakeholders of policy measures. Section 6 summarizes the results and lessons of these innovations; for performance auditors themselves, but also how they should consider monitoring and evaluation.

Performance audit, norms and criteria: Sense through simplicity

Performance audit can be defined as ‘the independent examination of the efficiency and effectiveness of government organizations, operations, or policies, with due regard to economy’ (INTOSAI, 2010: 11). The aim of performance audit is that it should lead to improvements (2010: 11).

As this definition by the organization of supreme national audit institutions INTOSAI shows, ‘performance’ has a broad meaning. Performance audit may appraise efficiency and effectiveness by investigating the output and/or the outcome of organizations, activities, and policy programmes. An illustration of this is a recent audit by the Netherlands Court of Audit of the effectiveness of various instruments to stimulate energy savings (Rekenkamer, 2011). The add-on in INTOSAI’s definition ‘with due regard to economy’ betrays performance audit’s connection to financial and operational auditing. Performance auditors may, for example, assess the costs of certain products and services while pointing out particular cost-drivers. They may compare the costs and quality of an agency’s output against a benchmark (INTOSAI, 2010). Performance audit can also include the examination of information systems, performance measures, and monitoring and evaluation arrangements (INTOSAI, 2010).

Traditionally, performance audit is a compliance-oriented craft (see Pollitt et al., 1999). Performance auditors work on the basis of professional audit norms and criteria that constitute the basis for performance judgements (Reichborn-Kjennerud and Johnsen, 2011). When assessing an organization’s output, the annual ‘production targets’ or the objectives to ‘improve quality’ or ‘reduce costs’ constitute logical starting points. When assessing the performance of policy programmes, norms and criteria can be found in policy objectives and indicators (e.g. the ambition to

reduce the number of traffic casualties or to improve water quality). INTOSAI guidelines stress that performance audit is based on the principles of good governance and decisions made or goals established by the legislature (as expressed in laws and regulations). Policy objectives and indicators – and, hence, norms and criteria – can also be derived from the executive (policy documents).

The emphasis on norms, criteria, and responsibilities is reflected in typical areas that are of interest to performance auditors: is there a clear structure of performance goals? Have the appropriate priorities and instruments been chosen for the use of public funds? Is there a clear distribution of responsibility between different levels of authority? Is there general cost awareness? Is there an adequate emphasis on management controls and reporting requirements? (INTOSAI, 2010: 12). All these questions relate to clarity performance auditors expect: there should be a clear structure, clear responsibilities, cost awareness, adequate emphasis on governance, etc.

Taken together, these norms and criteria, and especially the policy objectives and indicators constitute a relatively stable frame of reference. For politicians, policymakers, and performance auditors alike, this has important advantages: it provides focus, transparency, and predictability. It makes things simple and allows for clear statements, attractive political campaigns, and a clear division between policy and implementation. The active use of policy objectives, goals and indicators provides focus. They help to reduce complexity in thought and debate, which enables policy makers to develop a clear, communicable vision on how to achieve what. Objectives and indicators form easy-to-understand frames of reference in a policy-oriented debate (Van der Knaap, 2006). To sum up, paraphrasing Philips' advertisement slogan: the use of norms, criteria, policy objectives and indicators allows for sense through simplicity.

They fit well, too, into the rationality of the results-oriented, new public management paradigm, in which expectations about policy measures and their effects set the stage for managing, measuring, and judging efficiency and effectiveness (see Bouckaert and Halligan, 2008). When agreement is reached on objectives and indicators, they form the starting point of collective action on the basis of which interim adjustments may be made and learning can take place. In addition, clear norms and criteria, together with sound quality controls in the audit process, helps to produce credible reports on government performance (see IIA, 2006).

Complexity and dynamism: The call for responsiveness in evaluation and audit literatures

It may be naturally tempting to think of society's problems as simple, 'controllable' or 'tame-able' and of policy as a 'sequence' of pre-planned measures. And yet, as we experience every day: we live in a complex and ever-changing world – circumstances, knowledge and preferences will differ and continue to change (Room, 2011). One of the main pitfalls of performance audit that focuses firmly on objectives, targets, and criteria is that the auditor misses the true explanation of disappointing agency or programme performance by (1) continuing to stress the importance of objectives, targets and criteria, (2) criticise shortcomings in this domain, without (3) reflecting on the questions whether those performance indicators actually relate to society's real problems and stakeholders' concerns (cf. WRR, 2004, 2007; for a case description, see Reichborn-Kjennerud and Johnsen, 2011).

There is a classic episode of 'Yes Minister' in which Sir Humphrey meets for a drink with Sir Ian Whitchurch, Permanent Secretary at the Department of Health and Social Security. The latter expresses surprise at Hacker's concern over St Edward's Hospital. Whitchurch argues that one can't set up a hospital and start it running smoothly when there are patients around. The case for the

patient-free hospital is further strengthened by its superb performance on staff satisfaction and overall quality, measured by the absence of accidents, and complaints. I hope the point is clear: by focusing too much on the stable frame of reference, the auditor – like Whitchurch – may miss the entire point.

Although my example is, of course, rather ridiculous, the concern that performance audit might become less relevant is widely acknowledged (Power, 1997; Van Thiel and Leeuw, 2002; Bouckaert and Halligan, 2008; Lonsdale et al., 2011). It becomes increasingly clear that while the ‘objective-and-indicators-driven’ approach brings advantages, it also carries risks. Objectives, targets, and criteria may lead to a misunderstanding of other people’s needs, resistance to change, and unwanted strategic behaviour or gaming (see Van der Knaap, 2004, 2006). They are – by nature – logical abstractions of reality and represent a simplified version of reality. A narrowed interpretation of audit criteria may limit the questions evaluators and auditors can answer (Reichborn-Kjennerud and Johnsen, 2011; see also Greene, 1999).

Here, there may be a clear difference between ‘typical evaluators’ and ‘typical auditors’. As Ling puts it, evaluators are able to live with ‘multiple truths’ (2007: 133). They may be more inclined to accept the notion that circumstances may change and the different stakeholders may have different views on the performance of a programme at different times. While auditors often take a different view even in audit literature, due attention is given to responsiveness and the need to recognize the complexity of public sector interventions (e.g. Bemelmans-Videc et al., 2007; De Lancer Julnes et al., 2007; Ling and Wilkins, 2011).

As Bemelmans-Videc et al. (2007: 248) note: ‘Rarely does a program or intervention work completely independently of the interventions of other actors or the influence of other factors’. In fact, many explanations for disappointing policy results point to a ‘gap’ between simple policy intentions and schemes on the one hand, and the complex and ever-changing nature of societal problems on the other (2007: 248). The authors argue that in changing environments, ‘responsive programs *should* be changing their objectives to take advantage of opportunities or to respond to feedback [including feedback from audit and evaluation]’ (2007: 249). They envision a new type of accountability in which public sector auditing should play its role by (1) taking a dynamic rather than a static approach, (2) focusing primarily on results rather than on process, (3) stressing the need for continuous and responsive learning (2007: 253).

As the readers of this journal will testify, the risks of the new public management paradigm and its focus on all things measurable are also widely discussed in the policy evaluation discourse. Authors like Guba and Lincoln (1989), Schwandt (2002), and Hoppe (2010) discuss the ever increasing complexity of modern society’s problems. According to Schwandt (2002: 9): ‘One normative ideal is that monitoring systems ought to replace the complex social–political processes entailed in the design and delivery of social and educational services.’ His argument is clear: performance indicators can never do that and will even distort the understanding that we develop. It is likely that the use of objectives, targets, and indicators may result in not recognizing unexpected effects or changes in circumstances or public preferences.

It is important to point out that these pitfalls are not necessarily remedied by evaluation. On the contrary: similar effects will occur if evaluations or performance audits are tailored too strictly to existing policy targets and performance indicators. In this respect, the 2004 evaluation of the budget reform in The Netherlands found that, in general, evaluation reports make no attempt to compare recent developments in policy priorities with long-term objectives and trends as formulated in policy agendas: ‘Questions such as “Are we on the right course?” are rarely asked’ (Ministry of Finance, 2004: 19). In addition, the basic question as to why government is involved at all in certain issues are not addressed by evaluations. According to the 2004 evaluation, there is a ‘bias towards conservatism’ (2004: 19).

To counter this, in its 'policy audits' the Ministry of Finance gives specific attention to the underlying problem analysis, the actual relevance of the policy objectives, and the reason for government intervention. Audit questions include questions like: what is the problem that gives (has given) cause to the policy? Is this problem still a topical subject? Why does central government (still) consider it their responsibility to solve the problem? What are important positive and negative side effects?

Evaluation literature has witnessed a strong call for a more responsive, argumentative approach. Herein, the evaluation would be geared less to formal objectives and indicators (policy intentions), and more to actual activities, attitudes, and 'real results'. In short, the evaluator should try to establish the actual effects of a policy programme, not just those adopted by a central actor (Scriven, 1991; Fischer, 1998). Various forms of responsive evaluation have emerged (e.g. 'empowerment evaluation' [Fetterman and Wandersman, 2007], evaluation as a 'narrative' [Abma, 2005] and 'realist evaluation' [Pawson and Tilley, 1997]). The main characteristics are (1) taking criteria for judging success from both policy makers and stakeholders, (2) deliberate attention to weaker parties, and (3) engaged, qualitative methods that rely on conversation and dialogue (Abma, 2005). When reporting on the results of policy programmes, the evaluator would stress the context-bound nature of findings while explicitly referring to stakeholders' opinions, values and standards, especially when they differ from official policy documents (Abma, 2005).

Following the responsive evaluation school, researchers become less determined or 'programmed' in their assessment of outcomes as 'success' or 'disappointing' by existing formal objectives and indicators. Instead, the researcher could be more responsive to actual changes in perceptions, attitude, and/or behaviour. As I see it, this kind of responsiveness to social change and changes in stakeholder's perspectives is crucial for any political system that wants to intervene in society's problems in an intelligent, informed, and well-balanced way (Van der Knaap, 2006). Especially when applied in a system of results-oriented budgeting, where policy objectives and performance indicators play important roles, the added value of policy evaluation should principally be the systematization of the possibility that dialogue will reveal differences (Majone, 1989).

Bringing responsiveness in performance audit: New international guidelines and recent initiatives by the NCA

The need for performance audit that allows for more complexity is also recognized by the organization of supreme national audit institutions INTOSAI. In its 2010 guidelines, there is the following warning:

Performance audit should not be streamlined. Advanced performance audit is complex investigatory work that requires flexibility, imagination and high levels of analytical skills. Streamlined procedures, methods and standards may in fact hamper the functioning and the progress of performance audit. Consequently, standards – as well as quality assurance systems – that are too detailed should be avoided. Progress and practices must be built on learning from experience. (INTOSAI, 2010: 29)

It should be noted that the new INTOSAI guidelines include auditing the 'actual impact' of activities compared with the intended impact as part of performance audit (INTOSAI, 2010); an orientation towards citizens' needs is advocated. Unlike compliance auditing, performance auditors should not just assess whether policy programmes are being carried out according to plan. Neither should they focus too narrowly on whether objectives are reached, criteria met, and indicators attained (2010).

The new guidelines resemble initiatives by the Netherlands Court of Audit (NCA). The NCA wants its performance auditors to take a more responsive approach and stress the need for learning. Auditors must be more responsive for what auditees and other stakeholders have to say about the relevance of audit questions, the meaning of its intermediate findings and the usefulness of potential recommendations (NCA, 2005). To the NCA, there are three dimensions to this:

selection of audit themes and topics: putting social relevance and risks central;

audit process and methods: adopting a more participatory approach and conducting ‘reality checks’ – compact audits of the connection between policy programs and stakeholders’ needs and preferences;

follow-up audits: keeping track of the use and effects of performance audits.

This section describes these more participatory approaches to the audit process and the conduction of follow-up audits. Special attention to ‘reality checks’ is given in Section 5. (For a discussion of a selection of audit themes, see Turksema and Van der Knaap, 2007.)

Audit process: Adopting a more participatory approach

The NCA has actively opted for a more open and responsive approach, including more participatory methods. Following the example of the UK’s NAO (National Audit Office), NCA has developed a ‘menu for effective performance audits’ (NCA, 2005). This menu presents different participatory audit methods and other stakeholder involvement activities that audit teams can use during the different stages of the audit process.

The participatory audit approach consists of two phases: (1) actor analysis; and (2) the selection of possible methods and actions. In the actor analysis phase, the NCA auditors set out to analyse the main policy actors and other stakeholders and their respective interests. The NCA auditors look into the main parties involved, their involvement in a policy programme, recent communications and possible perceptions on the upcoming audit. In addition, the sensitivity of the topic in question; the complexity of the policy domain or network; the level of agreement on problem definitions, relevant values and norms; and the necessity for improvement are all considered. If there is a relatively high level of trust between auditors and those ‘under scrutiny’, a more cooperative approach may be fruitful: lessons and insights may lead to early adoption by a ministry or stakeholders. If, on the other hand, interests differ and there seems to be little trust between parties, a more formal approach may be better. Here too, the inclusion of external experts and stakeholders may help to obtain specific insights or specialist feedback.

In the second phase, the audit team makes a selection of possible actions. Along the timeline of the audit process, actions like meetings at the start of the audit and brainstorming sessions to discuss intermediate outcomes and potential remedies are considered and plotted. An important part of this is expert-panels in which auditees and other stakeholders are invited to discuss intermediate audit findings and possible programme improvements. As an example, in the NCA’s audit of the energy savings policy programme, a panel session was organized with auditees from the Ministry of Economic Affairs and the Ministry for the Environment. In addition, representatives from the paper and plastics industry, research and university institutions, and an environmentalist awareness organization were invited. The NCA team presented their intermediate audit findings and discussed possible conclusions and recommendations. While the Board of NCA always has final responsibility

for audit conclusions and recommendations – a fact that is always pointed out to participants – this kind of session allows for an active ‘testing’ of audit findings.

Follow-up audits: Keeping track of the use and effects of performance audits

The NCA wants to know what happens with audit recommendations after audit reports have been published: were recommendations followed and did that really help? To this end, the NCA actively monitors the ministers' follow-up to its audit recommendations and whether they fulfil their undertakings over a longer period of time, if necessary for five years or longer. In 2006, for instance, the NCA published a report on jobless people without basic qualifications. The NCA investigated central government policy to give jobless people aged 23 and over without basic qualifications better prospects of finding a job. Three years later, in 2009, an impact assessment was carried out on the three main recommendations.

As the two questions ‘were recommendations followed?’ and ‘did that really help?’ indicate, the follow-up audits have two dimensions. The first dimension is rather straightforward: did the ministries do what they promised to do? From the example above, the NCA concluded that regional cooperation to help jobless people without basic qualifications had indeed been strengthened. Within the second dimension, the NCA investigates which – positive and negative – changes actually took place concerning the issue at stake. Here, much harder questions are asked: how did the follow-up of audit recommendations contribute to the quality of implementation processes? Were jobless people better helped? To what degree was the social problem solved: did they succeed in finding new employment?

In practice, as assessing the impact of performance audits is expensive, many follow-up audits tend to be limited to the first dimension (see Pollitt et al., 1999; Morin, 2001; Pollitt, 2006). Currently NCA is developing new, cost-effective mechanisms to connect changes in policy programmes to changes in programme performance, including societal effects. Making use of existing M&E systems and involving ministries and other stakeholders are likely to be part of it.

‘Reality checks’: Testing policy theories, objectives, and indicators

The latest and perhaps boldest initiative to develop a more responsive approach to performance audit are the NCA ‘reality checks’, 20 of which were carried out between 2009 to 2011 (see NCA, 2010, 2011). NCA deliberately used this rather provocative term: in the language of innovators and product developers, a reality check is a test to make sure an idea is consistent with the ‘real world’. A new police strategy, for instance, is often tried out on the streets to find out whether it really works.

The purpose of NCA’s initiative was to provide government and Parliament with information on the actual appreciation and effects of public policy programmes. Or, in the words of the NCA President when she presented the results of the first round of reality checks to Parliament: ‘With our reality checks the NCA wants to bring the perspective of citizens and businesses to the *Binnenhof* (the seat of central government in The Hague). For this reason, NCA started out by looking at a number of problems that affect citizens and businesses and the ways in which government translated these problems into policy objectives and measures. The pivotal question was: what is the actual contribution of policy measures to find solutions as judged (1) from the position of the policy target group (the persons, businesses, or institutions that are experiencing a particular set of problems) and (2) from available evidence of programme effectiveness.

The NCA used the following questions in its audits:

- How do these stakeholders relate to the objectives, criteria, and intervention logic or ‘policy theory’ of central government?
- Do they recognize the relevance and value of it and do they use it?
- How do they appraise the actual interventions or policy tools and the way the policy programme is implemented?
- To what degree did the targeted spending actually ‘land on target’?
- What concrete impact does the policy have for those directly concerned?
- Which information is provided by central government on the effectiveness of the policy measures taken?

NCA auditors interviewed representatives from target groups and other stakeholders to establish what they thought of official objectives and criteria. The actual use of policy schemes by target groups was also analysed, using available data from agencies, statistics, and monitoring and evaluation reports. Where suitable and possible, auditors observed real-life negotiations, meetings, transactions, etc. Finally, the results were put forward to the responsible policy makers for a reaction.

NCA reported its findings to Parliament in May 2010 and 2011. Most of the 20 policy programmes the NCA investigated were considered to be relevant and focusing on real problems by those on the ‘receiving end’ of measures. The objectives – intended outcomes or societal effects – are recognized as relevant and valid. Most people want less crime or more safety. The problem lies in the preconceptions policy makers appear to hold when choosing intervention schemes to achieve those objectives. At the level of policy tools, the checks revealed that the link between policy and practice leaves a great deal to be desired. Although expected output (grants, subsidies, websites, advice) were mostly delivered according to plan, recipients were either not very familiar with it, could not use it, saw no need for it, or expressed a greater need for other measures (NCA, 2010, 2011).

Example 1: Improving safety in small businesses

By means of a subsidy scheme small businesses are encouraged to take preventive measures against crime. The compensation of costs for ‘tailor-made advice’ on how to improve security is a key element in the policy programme. In practice, however, many hired security consultants and specialists work with standard formats. NCA found that although 80% of the target group found the safety scan useful, only 8% thought the scan was worth 350 euros.

Example 2: Energy saving measures for existing homes

There are three subsidies to improve the energy efficiency of existing homes. NCA found a clear tension between the objective to relieve participating homeowners of the administrative burdens of application and declaration of costs on the one hand, and the reluctance of those homeowners to leave this to contractors on the other.

Example 3: Web-based assistance for the chronically ill

Through the website ‘regelhulp.nl’ national government aims to provide assistance in applying for care, support or financial support for the chronically ill, disabled and elderly. The objective is to reduce the

administrative burden for these groups. NCA found that users of the site are generally positive about the usability, simplicity and clarity of the website but that they did not consider the applications themselves as the main administrative burden. Of the potential total target group of 3.3 million clients, the website's reach is limited to 1.2%.

As these examples show, the connection between top-down interventions on the one hand, and the motives and needs of target groups on the other, is less than optimal. The choice of instruments can be characterized as 'supply-side driven'. In quite a few cases, NCA came across a 'scatter-gun approach'. Instead of evidence-based interventions – where there would be a clear, well-established relationship between intervention and effect – often a rather rash choice of subsidies or other financial incentives was made. 'Money will work wonders' seems to be the dominant perspective, even when other factors (cultural, technical, judicial) are behind societal problems.

Example 4: Primary education: Rewarding good teachers

Doubt on whether it was real capability or mere seniority that led to actual promotions and pay rewards in primary schools prompted a central government Action Plan for Teachers. According to the plan, the best performing teachers would get promoted and should receive more pay. NCA found that schools that already have many teachers on higher salary scales, gave promotion to poor performing and inexperienced teachers instead. In addition, schools are reluctant to make longer-term commitments as there are legitimate concerns about the availability of future funds for schools. Several school managers gave evidence that the plan sometimes leads to less rather than more differentiation in teachers' pay.

Example 5: Children's participation in sporting activities

Policy programmes to increase the participation of children in sporting and other organized activities are carried out by the municipalities. Central government, which transfers the budgets to local government, has no say in the actual projects, but it can and does make agreements to stimulate municipalities to give families 'tangible help' (for example: paying for sporting club's membership fees). NCA found that money did not constitute the main barrier for parents to encourage their children to be active in sports. Instead, the explanation can often be found in different cultural or religious backgrounds.

The reality checks showed that it is important that policy makers – both prior to the introduction of policy measures and during their implementation – actively test out whether the assumptions behind the policies make sense. Good ex ante evaluation of available options and a thorough consideration into how policy instruments – including their design and/or implementation – can lead to the achievement of desired results that are often lacking. Partly as a result of poor policy preparation, NCA found a lack of relevant data and information to follow the implementation and results of policy interventions and make necessary corrections. Information on the relationship between policy measures and effects is rare. Instead, information is provided on outputs and what could be labelled at best 'intermediate effects' (e.g. the number of people 'reached' by a programme).

On the positive side, NCA auditors also found cases where government did try to map target groups' needs and preferences before embarking on new policy initiatives. Consultations, city hall meetings, and expert and citizens panels (including Web 2.0 applications) are examples of this. In addition, reports by central planning agencies and evaluation reports were sometimes used to tailor policy measures to the issues at stake and the concerns of those affected.

Example 6: Stimulating rail transportation

The policy to increase the number of miles travelled by passengers on the public rail system in The Netherlands was established after a broad social consultation. Elements mentioned by travellers were included in the programme (e.g. better signposting).

In many cases NCA found that the way policy interventions were designed led to rather complex implementation processes. There are often many links between government, intermediate organizations, stakeholders, and target groups. For both its ex ante information (what is needed) and the actual delivery of policy measures (what can be done), central government is often dependent on other organizations, notably local government and quasi autonomous governmental bodies ('quangos'). The risk of this interdependence is 'the problem of many hands' and long delivery chains. In its slipstream, institutional rationality may come to crowd out societal relevance, which may pose a real risk to both efficiency and effectiveness.

Towards sense and complexity: Conclusions, challenges and discussion

During the last decades, when many OECD countries witnessed a 'steady ascendancy of performance as a fixation of governments', much has been invested in the development of performance measurement and management (Bouckaert and Halligan, 2008). Parallel to this, performance audit has established itself firmly next to regularity and financial auditing. With achievement came criticism and the worry that too much focus on objectives and indicators will limit our understanding of reality. While making things simple and measurable is often a good thing, dealing with complexity and change remains an important challenge for performance management and performance audit alike (Mayne, 2006; Ling, 2007). We should continue to beware of congratulating hospitals that perform excellently when there are no patients around.

Performance auditors are aware of this risk. As the new INTOSAI guidelines and the initiatives described in this article demonstrate, they actively seek new ways to mitigate the negative consequences of their compliance-oriented craft.

The first results of the 'reality checks' by NCA are promising. They demonstrate how wrong assumptions – together with a tendency to employ complex, sometimes paternalistic instruments and implementation mechanisms – lead to a mismatch between policy measures and target groups. In addition, they uncover how much of the complexity in the relationships between central and local government, or in the implementation by networks of quangos and other administrative bodies, is in fact man-made. Moreover, the reality checks underline the importance of good ex ante evaluation: a thorough problem analysis is a prerequisite for successful public policies and their performance management.

Nonetheless, becoming more responsive is not an easy process. Recent initiatives by the NCA to bring more responsiveness into performance audits illustrate the challenges in finding the best of both worlds:

- First, there is the problem of timing. Early warnings in the form of 'reality checks' may prevent inefficient or ineffective programmes from going from bad to worse. Still, to ministries and agencies, early audits and evaluations may lead to the feeling that a programme is not given the opportunity to demonstrate its worth. Opposition and resistance can be expected and may lead to sharp discussions on methodological issues. Being explicit about the scope and representativeness of audits is of paramount importance. While not ruling out

concerns about timing and methods, user panels and social media present new tools for auditors, evaluators, and policy makers alike to learn from the insights and preferences of target groups and stakeholders (Dahler-Larsen et al., 2007).

- Second, responsiveness should not be mistaken for indiscriminately using stakeholders' concerns and complaints against auditees. It is relatively easy to find complaints and disparities between public policy programmes intentions and target groups' needs and wishes. Interventions may lack relevance in the eyes of respondents, but the question arises when exactly that should be called wrong. When a policy measure is criticised by a target group, this may point to a bad policy design; it may also stem from overconfident or otherwise biased respondents. For a well-balanced judgement it is necessary to examine available ex ante evaluations and the decision-making processes.
- Third, there should be fairness in criticising the use of policy objectives and performance indicators. The reality checks showed that the 'standard glove' of many policy initiatives did not fit the receiving hand. To a large degree, this can be blamed on an active simplification of problems to available solutions, to answer complex questions with easy answers, and to apply routine and standard formulae to different cases. However, objectives and criteria are and always will be deliberately simplified versions of (desired) realities. For both political debate and implementation practice, this has advantages as well as drawbacks. So at this point too, an examination of available ex ante evaluations is important for a fair judgement.

Responsive performance audit does not mean that auditors should abandon their traditional criteria for judgement or become a mere passive receptor of other people's interpretations and criteria. Neither should performance auditors join the lamenting choir that denies all benefits from policy objectives, performance targets, and indicators. It is true that the use of norms and criteria may lead to limited evaluation and audit questions and, as a result, limited answers. This indeed may decrease the value of these evaluative activities: the risk is that if people are happy to monitor and evaluate performance against inappropriate or limited norms, they may not see failure or find the explanations for disappointing success (see Van der Knaap, 1995, 2004). For this reason, performance audit should look for sense *and* complexity. While maintaining independence and benefiting from the focus that the simplicity of norms and criteria provide, it should also be open and responsive to complexity and change.

Diversity in values and preferences, complexity, interdependence between stakeholders, and change will always be with us. By actively seeking insights and experiences from stakeholders and target groups, policy makers, evaluators and auditors alike may help to provide feedback on the validity of existing policy schemes.

In a democratic society, government has a moral duty to organize its own responsiveness and reflexivity. With the rise of the internet and new social media, there are now more possibilities than ever to engage the broader public to policy making and evaluation (e.g. recovery.gov; Dahler-Larsen et al., 2007; In 't Veld, 2010). Learning and the organization of policy-oriented learning are crucial in this respect. How are governments and agencies trying to find out what will work best in which context? How do they consult with stakeholders? How do they organize their own feedback and reflection on existing policies, objectives, and criteria through monitoring and evaluation? And, crucially: how do they translate lessons into better practice?

In answering these questions, performance audit has an important role to play. It can help to balance the benefits and risks of policy objectives and performance indicators by looking critically to existing monitoring and evaluation practices: how are these instruments used for responsiveness and policy-oriented learning?

People want effective policy programmes. In this never-ending quest, politicians, policy makers, evaluators, and performance auditors alike should be well aware of the ‘taming trap’ that well-articulated objectives, indicators and criteria bring with them. Problems appear ‘under control’, programmes and their implementation successful, while important negative consequences may be overlooked. This awareness may indeed call for an abandonment of what American psychiatrist Morgan Scott Peck called the ‘urge to simplify everything, to look for formulas and easy answers’. We must not try to simplify everything, although we should keep simple what already is simple. In addition, we must at times – when problems are very serious, try to come to shared definitions of society’s problems and consensus on how to tackle them. At both points, responsive forms of evaluative inquiry can make valuable contributions.

Note

The author would like to thank Gert Jan de Vries, Jeroen Kerseboom and Rudi Turksema, and two anonymous referees of this journal for their comments on an earlier draft of this article.

References

- Abma TA (2005) Responsive evaluation: its meaning and special contribution to health promotion. *Evaluation and Program Planning* 28: 279–89.
- Bemelmans-Videc M-L, Lonsdale J and Perrin B (eds) (2007) *Making Accountability Work – Dilemmas for Evaluation and for Audit*. New Brunswick, NJ: Transaction Publishers.
- Bouckaert G and Halligan J (2008) *Managing Performance – International Comparisons*. London: Routledge.
- Court of Audit (2011) Staat van de Rijksverantwoording 2010 (State of Government Accountability 2010), The Hague.
- Dahler-Larsen P, Breul JD and Boyle R (2007) *Open to the Public – Evaluation in the Public Arena*. New Brunswick, NJ: Transaction Publishers.
- De Lancer Julnes P, Stokes Barry F, Aristigueta MP and Yang K (2007) *International Handbook of Practice-Based Performance Management*. Thousand Oaks, CA: SAGE.
- Fetterman D and Wandersman A (2007) Empowerment evaluation: yesterday, today, and tomorrow. *American Journal of Evaluation* 28(2): 179–98.
- Fischer F (1998) Beyond empiricism: policy inquiry in post positivist perspective. *Policy Studies Journal* 26: 129–46.
- Greene J (1999) The inequality of performance measurements. *Evaluation* 5(2): 160–72.
- Guba EG and Lincoln YS (1989) *Fourth Generation Evaluation*. Newbury Park, CA: SAGE.
- Hoppe R (2010) *The Governance of Problems – Puzzling, Powering, Participation*. Bristol: Policy Press.
- In ‘t Veld RJ (2010), *Knowledge Democracy*. Berlin: Springer Verlag.
- IIA (Institute of Internal Auditors) (2006) *The Role of Auditing in Public Sector Governance*. Altamonte Springs, FL: IIA.
- INTOSAI (2010) *Implementation Guidelines for Performance Audit – Standards and Guidelines for Performance Audit Based on INTOSAI’s Auditing Standards and Practical Experience*. INOSAI. URL: <http://www.intosai.org/blueline/upload/1implgperfaude.pdf>
- Ling T (2007) New wine in old bottles? In: Bemelmans-Videc M-L, Lonsdale J and Perrin B (eds) *Making Accountability Work – Dilemmas for Evaluation and for Audit*. New Brunswick, NJ: Transaction Publishers.
- Lonsdale J, Ling T and Wilkins P (2011) *Performance Auditing – Contributing to Accountability in Democratic Government*. Cheltenham: Edward Elgar Publishing.
- Majone G (1989) *Evidence, Argument, & Persuasion in the Policy Process*. New Haven, CT: Yale University Press.
- Mayne J (2006) Audit and evaluation in public management: challenges, reforms, and different roles. *The Canadian Journal of Program Evaluation* 21(1): 11–45.
- Ministry of Finance (2004) *Evaluation of VBTB*. The Hague.

- Morin D (2001) Influence of value for money audit on public administrations: looking beyond appearances. *Financial Accountability & Management* 17(2): 99–118.
- Netherlands Court of Audit (2005) *Essentials of Good Government*. The Hague.
- Netherlands Court of Audit (2010) *Staat van de beleidsinformatie 2010*. The Hague.
- Netherlands Court of Audit (2011). *Staat van de Rijksverantwoording 2011*. The Hague.
- Pawson R and Tilley N (1997) *Realistic Evaluation*. London: SAGE.
- Pollitt C (2006) Performance information for democracy – the missing link? *Evaluation* 12(1): 38–55.
- Pollitt C, Girre X, Lonsdale J, Mul R, Summa H and Waerness M (1999) *Performance or Compliance? – Performance Audit and Public Management in Five Countries*. Oxford: Oxford University Press.
- Power M (1997) *The Audit Society: Rituals of Verification*. Oxford: Oxford University Press.
- Reichborn-Kjennerud K and Johnsen Å (2011) Auditors' understanding of evidence: A performance audit of an urban development programme: *Evaluation* 17(3): 217–231.
- Room G (2011) *Complexity, Institutions and Public Policy – Agile Decision-making in a Turbulent World*. Cheltenham: Edward Elgar.
- Schwandt TA (2002) Democratization, globalization and performativity: implications for learning in and from evaluation. Keynote address given at the 5th biennial meeting of the European Evaluation Society, 12 October, Seville.
- Scriven M (1991) *Evaluation Thesaurus*. Newbury Park, CA: SAGE.
- Stern E (2011) Editorial. *Evaluation* 17(3): 213–15.
- Turksema R and Van der Knaap P (2007) Planning and selecting performance audits at the Netherlands Court of Audit. *International Journal of Government Auditing*, July. URL: <http://www.intosajournal.org/technicalarticles/technicaljul2007b.html>
- Van der Knaap P (1995) Policy evaluation and learning: feedback, enlightenment or argumentation? *Evaluation* 1(2): 189–216.
- Van der Knaap P (2004) Theory-based evaluation and learning: possibilities and challenges. *Evaluation* 10(1): 16–34.
- Van der Knaap P (2006) Responsive evaluation and performance management – overcoming the downsides of policy objectives and performance indicators. *Evaluation* 12(3): 278–93.
- Van Thiel S and Leeuw FL (2002) The performance paradox in the public sector. *Public Performance and Management Review* 25(3): 267–81.
- WRR (Wetenschappelijke Raad voor het Regeringsbeleid) (2004) *Bewijzen van goede dienstverlening*. Amsterdam: Amsterdam University Press.
- WRR (Wetenschappelijke Raad voor het Regeringsbeleid) (2007) *Lerende overheid – Een pleidooi voor probleemgerichte politiek*. Amsterdam: Amsterdam University Press.

Peter van der Knaap is Director for Performance Audit with the Netherlands Court of Audit. Prior to this he was Head of Policy Evaluation in the Netherlands Ministry of Finance. He has a PhD in Social Science from Erasmus University, Rotterdam. He publishes on policy evaluation, organizational learning, and the risks and challenges of performance management. Please address correspondence to: Netherlands Court of Audit, Directorate for Performance Audit, Lange Voorhout 8, NL-2514 ED, The Hague, The Netherlands. [email: peter.vanderknaap@ziggo.nl; website: www.petervanderknaap.com]