



Responsive Evaluation and Performance Management

Overcoming the Downsides of Policy Objectives and Performance Indicators

PETER VAN DER KNAAP
Netherlands Court of Audit, The Hague

Worldwide, governments are under increasing pressure to deliver results. There is general recognition of the importance of performance measurement and a results-oriented focus for effective public management. The shift from inputs to outputs and outcomes is accompanied by an increased use of performance indicators and policy targets. Critics point to the negative effects of static forms of performance-measuring and governance: simplification, resistance to change and a strategic use of indicators. They stress the need for an approach to governance that respects and is responsive to diversity and dynamism. Such an approach to public policy should recognize the importance of stakeholders' opinions and motives and, hence, dialogue. Herein, policy evaluation can help by systematically questioning the validity of policy goals and performance indicators. Based upon recent experiences with results-oriented budgeting in the Netherlands, this article outlines a possible approach to a responsive, 'verifying' policy evaluation. Against the background of the traditional functions of evaluation in results-oriented management (assessing efficiency and effectiveness), the paradoxical nature of performance indicators is considered: on the one hand they are 'frozen ambitions', yet on the other they must facilitate dialogue and learning. Building on these attributes, a dynamic perspective on evaluation is put forward. In this, using information and insights from stakeholders, policy evaluation should explicitly question the validity of the policy objectives and performance indicators *without* compromising their value and significance altogether. In this way, the evaluator may fulfil the ambition of the 'argumentative turn' in policy evaluation: to help to raise the quality of dialogue and decision-making between stakeholders.

KEYWORDS: accountability; performance management; policy-oriented learning; results-oriented budgeting; responsive evaluation

The Rise of Performance Measurement and Results-Oriented Budgeting

Although developments differ from country to country, performance indicators and policy targets are increasingly being used in policy documents and budgets to indicate what performance is expected, for what purpose actions are taken and at what cost (Perrin, 2006). According to a 2003 OECD/World Bank survey of budget practice, 32 percent of OECD member countries include non-financial performance data in their budget documents (OECD, 2004). For nearly 27 percent of OECD member countries, the inclusion of performance targets on government policy outcomes and/or outputs in the budget documents constitutes a legal requirement (OECD, 2004).¹

To learn more about the relationship between results-oriented budgeting and policy evaluation, this article considers the case of the Netherlands in more depth. In the Netherlands, financial regularity has improved greatly since the 1980s (Sorber, 1999). In the 1990s, the focus shifted from legitimacy and regularity to efficiency and effectiveness. The main driving force behind this was a group of parliamentarians who wanted a more transparent and policy-oriented debate on government budgets and annual accounts. This led to close cooperation between Parliament, the Ministry of Finance and the Court of Auditors to launch an initiative to promote efficient and successful government action through a broad implementation of the results-oriented management, budgeting and accountability model. The intention was to improve government management and performance by clarifying the relationship between the deployment of resources, products and services and the effects these aim to attain, as well as to take this as a starting point for policy-making, implementation and political debate.

In 1999, under the title 'From Policy Budget to Policy Accountability' the Dutch government launched an initiative to improve the information value and accessibility of budget documents and annual accounts (Ministry of Finance, 1999; see also Van der Knaap, 2000, for an extensive overview). The general purpose of this so-called 'VBTB-project' was to make budget documents and the budgetary process more policy-oriented. The new-style government budget should answer three simple questions:

- What do we want to achieve?
- What will we do to achieve it?
- How much will our efforts cost?

At the end of the budgetary year, the government's annual accounts have to answer related questions, i.e.:

- Have we achieved what we intended?
- Have we done what we should have done in achieving it?
- Did it cost what we expected?

The general purpose of the 1999 proposals was to make the budgetary and accounting processes more results-oriented by systematically presenting information about intended and achieved policy objectives, policy measures or

Evaluation 12(3)

instruments, and their costs. To achieve this the budget is no longer organized into categories of financial expenditure. Instead, it is based upon the main policy objectives – for example, ‘the provision of a safe and affordable public transport system’.

Ever since its introduction, the effects of the results-oriented budgeting system have been monitored by both the government and the Court of Auditors. Table 1 gives an overview of the degree to which the three key questions of budgets and annual reports were answered.

Table 1. Percentage of Line Items Adequately Exemplified with Performance Indicators

	Budget 2003	Account 2003	Budget 2004	Account 2004	Budget 2005	Budget 2006
What do we want to achieve? Have we achieved what we intended?	34	29	35	27	41	53
What will we do to achieve it? Have we done what we should have done in achieving it?	51	46	63	55	72	66
How much will our efforts cost? Did it cost what we expected?	74	70	81	70	87	86

As Table 1 indicates, ministries are increasingly successful in including information on intended policy effects and measures in the annual budgets and accounts. Five years after the introduction of the results-oriented budget system, many of the government’s central policy objectives are specified, target groups of policy are stated, and the time frames are clear. The objectives are also explained by describing the direction of the effects to be achieved, and in many cases these are accompanied by performance indicators (Court of Auditors, 2005; Ministry of Finance, 2004: 11).

Traditional Functions of Evaluation Research in a Results-Oriented Budgeting System

In a results-oriented budgeting and accountability system evaluation research should complement financial statements, audit reports, operational management statements in budget accounts and policy reports by providing an insight into the *efficiency* and, if possible, the *effectiveness* of policy implementation (Ministry of Finance, 2001). Its purpose is to enhance the information derived from monitoring systems (e.g. how many young offenders are waiting to be treated or have been treated in an institution?) by helping to assess the degree to which policy measures made a difference (e.g. treatment in relation to the reduction in criminal behaviour).

In a results-oriented budgeting system, the results of evaluation studies are being – or should be – included in the annual reports. Together with performance indicators, information on goal achievement and explanatory texts, they constitute the ‘body of evidence’ that can be used to hold government accountable. The cycle is completed when government accounts for the use of resources and the results achieved and undertakes to make any necessary improvements. This is often referred to as the ‘accountability function’ of policy evaluation (Schwartz and Main, 2005).

‘Any necessary improvements’ bring us to the second function of policy evaluation: the ambition of having a ‘learning government’, i.e. a government that is capable of improving its policy measures and underlying assumptions or policy theories. When linked with ‘learning’, it becomes clear that evaluation is a normative activity. The key question for ex-post evaluations is: has the government chosen the right activities, has it carried them out well and did it achieve the desired results? Whether the government has chosen the right activities can be determined on the basis of the proven effectiveness of policy measures and, of course, on the basis of political convictions and preferences. The second dimension of evaluation relates to the implementation process: has the government organized its operations efficiently and is it conducting its policy well? When policy programmes perform unsatisfactorily, a critical review of underlying assumptions and hypotheses must then follow. This may lead to a refinement of the policy theories or the replacement or termination of the policy altogether.

Policy evaluation then contributes to rendering account and supports learning on the basis of high-quality information on government efficiency and policy effectiveness. At first sight, these traditional functions of policy evaluation seem still alive and well today. The central theme of Accountability Day in the Netherlands, for instance, is the systematic assessment of the original ambitions: were they fulfilled in an effective and efficient manner? Not by chance, the same questions are prominent in the evaluation chapter of the 2001 Government Account Act:

The main purpose of evaluation research is to provide credible insight into the following questions:

Has government policy achieved the desired results and to what degree can we attribute the results to policy measures? (i.e. policy effects and effectiveness); and

How well did government carry out its policies? (i.e. efficiency and other quality criteria, such as speed and accuracy). (Ministry of Finance, 2001)

The new regulations have now been implemented by all government departments. In June 2006, the central database for policy evaluations, the State Evaluation Overview (EOR), contained a total of 1681 evaluations in central government, and all are linked to the policy objectives of the new style budgets (<http://intra.ryx.nl/sw/rsw/eor>; accessed 23 June 2006).

It will be clear that most policy evaluations are positioned at the end of the policy process. With the exception of ex-ante evaluation during the policy preparation stage, accountability debates and learning from evaluation research seem typically to be ex-post considerations. Feedback at the end of the policy

cycle is considered to be the most important aspect of evaluation. There are, however, clear indications that this rather static use of evaluation will not be sufficient to meet the needs of an ever-changing and complex world, especially when the potential downside of performance indicators and formal policy targets is taken into account. But let's look at the benefits first.

Benefits of Policy Objectives and Performance Indicators: Providing Focus and a 'Working Language'

The use of policy objectives and performance indicators in policy-making and policy-oriented debates has two important benefits: it gives focus and it provides a language for policy, management and even political debates.

Attention and Learning: Focus

As the three questions of the *VBTB* project indicate, results-oriented budgeting strongly follows a rational-analytic approach to public policy. The answers to the three questions constitute a 'policy theory'. Such a theory expresses an expected causality between means, instruments and objectives: 'If we undertake these actions, good (intended) consequences rather than bad (unintended) ones actually will come about' (Wildavsky, 1987: 35).

Policy targets and performance indicators make theories concrete. Every such well-articulated policy theory, in particular one that is well-accepted, directs perception, interpretation and, hence, assessment. Schematic knowledge and assumptions embodied in theories, including those on public policy programmes, serve as the basis of our understanding of the world. Weick probably best describes the function of theoretic schemes: 'A schema is an abridged, generalized, corrigible organization of experience that serves as an initial frame of reference for action and perception' (Weick, 1979: 50). Policy theories, objectives and performance indicators fit well within this definition: within the frames of reference (or locus), they provide focus.

Accountability and Dialogue: Making Complexity Discussable

Government relies on other societal actors to achieve success. Cooperation, coordination and consultation are key elements of good governance. Formal policy targets and performance indicators form a kind of 'working language'. Making the assumptions behind policy measures concrete by using targets and performance indicators will help to facilitate political, policy and management debates. In a world that seems to become more dynamic and complicated all the time, it is often useful to temporarily and partially freeze ambitions in terms of performance and effect targets.

In a system of results-oriented budgeting and accountability, being answerable to the question 'did you do what you promised to do?' is fundamental. In this respect, it is only logical that performance indicators are deployed as 'yardsticks' to address and discuss the efficiency and effectiveness of government policies.

The evaluation of the *VBTB* project in 2004 reflects this logic: the presence and use of performance indicators has helped to draw attention to the efficiency and effectiveness of policy measures (Ministry of Finance, 2004). A change in the nature of questions, amendments and motions in the House of Representatives constitutes further proof (Ministry of Finance, 2004).

Risks of Policy Objectives and Performance Indicators: Simplicity, Resistance to Change and Gaming

As well as advantages, performance indicators and policy targets also carry risks. Authors like Power (1997) and De Bruijn (2001) repeatedly point to the risks of simplification, the tendency to develop tunnel vision and a general resistance to change, and a strategic use of targets and indicators.

Simplicity

Policy theories, targets and performance indicators are by their nature logical abstractions of reality. By definition, they present a simplified version of reality. Schwandt persuasively formulates the risk of simplification as follows: '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' (Schwandt, 2002: 9). His argument is clear: performance indicators can never do that and will instead distort the understanding that we develop.

In this respect, the 2004 evaluation of the *VBTB* project found that the abundance of output indicators sometimes tends to distract the focus of practitioners and parliamentarians from the *actual* policy objective:

Many of the objectives stated in the policy articles have to do with output or resources, such as 'euros spent on TV spots', 'acres of land', 'carrying out strategic research', 'conducting policy evaluations', 'the armed forces must be able to make a high quality, high-tech contribution to international operations ...', 'granting subsidies'. (Ministry of Finance, 2004: 45)

The conclusion is that, in those instance, performance indicators seem to have no bearing on the results of policy, or that they are completely off the mark:

The big drawback about performance indicators, key figures, performance contracts, etc., is that numbers start to determine policy. Among the unintended effects, reality may be represented too simply or management may be driven by inappropriate performance (the number of fines does not say anything about how safe the country is). The danger inherent in performance indicators is that it can lead to quantifying policy that can better be assessed in qualitative terms. This entails a risk of an unnatural effect of performance measurement. We must safeguard against this 'measurability paradox'. (Ministry of Finance, 2004: 45)

It is important to point out that this practice is *not* corrected by evaluations. Many of the evaluations carried out in the Netherlands between 2001 and 2004 tended to concentrate on the adopted policy theory and performance indicators. The 2004

Evaluation 12(3)

evaluation concluded that the majority of evaluation research was not aimed at the effects of policy. Usually the studies involved solely process evaluation or monitoring. Not a single study ‘managed to chart the effects of policy’ (Ministry of Finance, 2004: 33).

Resistance to Change

Traditional evaluation is usually based upon the following pattern of thought:

- Policy, once decided, should be carried out.
- Later changes are irrelevant as long as they are not adopted as policy changes.
- The only good evaluation is an evaluation that analyses the costs and benefits of existing policy measures. (De Bock et al., 1996)

As a consequence, policy targets and performance indicators may have a ‘blinding’ effect. There is a risk that the development of targets and indicators and the ongoing monitoring of implementation and results demand so much attention that unexpected effects or changing public preferences are simply overlooked (Gerritsen and Geut, 1997). A similar effect can occur if policy evaluations are too strictly tailored to existing policy targets and/or performance indicators: the evaluation research focuses exclusively upon the fixed policy and theory and its associated goals and performance data.

In this respect, the 2004 evaluation of the budget reform found that:

Generally, the policy report makes no attempt to compare recent developments in policy priorities with long-term objectives and trends as formulated in the policy agenda. Questions such as ‘Are we on the right course?’ are hardly ever asked. (Ministry of Finance, 2004: 19)

In addition, the question about government funding is not asked explicitly: why is the government involved in solving this problem in the first place? These practices, too, are not corrected by evaluations. This bias towards ‘conservatism’ may be explained by the fact that there is doubt with respect to the independence of the way evaluations are commissioned and/or carried out (Ministry of Finance, 2004: 19).

Gaming

Policy implementation is often a long process. Also, in many cases, government is not the dominant actor – and even if it is dominant, the government is not omnipotent. The combination of these factors means that the reflexivity or learning capability of elements being ‘governed’ (e.g. those guilty of tax fraud or speeding, but also university boards and even other government agencies looking for ways to optimize their funds) has every opportunity to erode the effectiveness of policy measures. Especially when applied in ‘steering relationships’, strategic behaviour will inevitably lead to learning processes and strategic behavioural responses that will gradually erode the value of performance indicators (In ‘t Veld, 1989: 28).

Their utility will decline more rapidly as individual interests increase and the performance indicators' various functions conflict with each other.

Experience teaches that one-sided steering/control towards results tends to lead to rather perverse effects (Scientific Council, 2005: 3). This will be especially problematic where performance indicators are initially developed to facilitate a learning process on policy options and/or potential improvements to policy measures, but become used primarily in accountability systems. The risk is that this will lead to direct conflicts and very probably to information no longer being supplied – or supplied in a biased way – by agents and other stakeholders.

The strategic dimension of using performance indicators can also be discerned in results-oriented budgeting. By presenting a large number of policy targets and performance indicators some ministries may effectively conceal the real issues at stake in discussing the merits of policies or policy implementation. In some policy domains, especially those with many subsidized measures, the phenomenon of 'data dumping' leads to bulky and often incomprehensible documents. For example:

True insight into the key questions is still too often frustrated by the natural tendency of officials and administrators to hedge their bets and give veiled answers. After all, it saves them problems. But it is definitely an obstacle to the Lower House doing its work properly. (Ministry of Finance, 2004: 17)

Bovens and 't Hart depicts this phenomenon as the 'blame game': strategies to minimize or avoid blame in case of failures and to maximize credits for successes (2004: 7, after Hood, 2002). Because of this gaming, authors like Behn and Power point to the dilemma between accountability and efficient government. Too much emphasis on accountability – and, therein, the importance of policy targets and performance indicators – may lead to 'sub-optimal and inefficient decisions instead of improved performance' (Bovens and 't Hart 2004: 8). In the Netherlands, this risk was stressed by an influential report by the government's Scientific Council on the 'Proofs of good service provisions'. It argued that focusing exclusively on measurable performance can lead to a destruction of knowledge, a 'dictatorship of mediocrity' and a 'stacking-up of supervision and responsibility' (Scientific Council, 2005: 4).

The Argumentative Turn Revisited: Systematically Incorporating Responsiveness, Critical Reflection and Debate in Results-Oriented Budgeting

As has been argued, policy theories and related performance indicators and targets are abstractions of reality (for an overview, see Van der Knaap, 2004). Freezing the desired performance and societal outcome in indicators helps us to understand the world in an efficient manner but may also distort our worldview. Neither policy nor the relationship between government and society (or between government organizations themselves) is static. Carving policy objectives and performance indicators in stone does no justice to the dynamic nature of many

policy processes. As the OECD rightly states in the synopsis of *Government of the Future*,

The public's needs are rapidly changing as societies become more diverse, complex, and fragmented. The pace is faster than ever: governments cannot rely on one fixed set of solutions, but need to listen to the ever-changing demands and innovate to find solutions. (OECD, 2001: 2)

Yet if ambitions are not frozen, no reference point is available for a dialogue: if everything is fluid, nothing is certain (it cannot even be discussed). The challenge is not to shy away from freezing but to be constantly willing 'to let certainties unfreeze'. This challenge is closely linked to the paradox of static and dynamic quality as described by Pirsig (1992; cf. the 'static logic' and 'dynamic creativity' of De Wit et al., 2000). As I see it, the continuous interplay between freezing and unfreezing, between static/analytical logic and dynamic/responsive creativity, is part of the pursuit of any intelligent set of policy measures. The paradox is that the meaningful application of performance indicators stands or falls on the ability to accept them and to put them into perspective. It is precisely in this area that policy evaluation – taking up more argumentative, responsive functions – can help.

As said, when the results-oriented budgeting and management system was introduced in the Netherlands, evaluation was conceived primarily as an instrument to produce high-quality information on government efficiency and policy effectiveness. This employment neatly fits the traditional rational-analytic approach to policy-making and evaluation (Dunn, 1994).

Since the late 1980s, the 'traditional', rational-objectivist model of policy evaluation has been seriously challenged (Dryzek, 1982; Everitt, 1996; Fisher and Forester, 1993; Majone, 1989; see also Van der Knaap, 1995). Instead, an argumentative-responsive approach has been put forward, in which evaluation:

- remains sceptical of rational-analytical techniques and modes of practice;
- is geared more to actual activities, performance and attitudes than to formal objectives and indicators (policy intentions);
- recognizes the importance of moral debate and respects the right of every stakeholder to legitimate opinions;
- (by letting go of objectives and a central actor perspective) is relatively value free in its assessment of such outcomes as 'success' or 'disappointment';
- tries to respond to the information needs of administrators, stakeholders and the general public; and
- when reporting on the success and failure of a policy programme, explicitly refers to the stakeholders' different values and standards (after Stake, in Abma, 1996: 63; Everitt, 1996: 180; see also Stake, 2004).

In addition, authors such as Guba and Lincoln (1989), Abma (1996) and Schwandt (2001) stress the pluralist, interdependent nature of both government and society. First, the development and implementation of policies requires the support, participation or even cooperation of many actors. Second, the complexity

of modern society's problems appears to be increasing every year. No longer can policy measures be rationalized solely by means of a stern, efficient and rational central governmental, that, on the basis of sophisticated research, allocates resources and decides on binding values and norms with authority (In 't Veld, 1989). In sum, central to the argumentative-responsive approach is the belief that, through constructive argumentation, policy actors, networks or advocacy coalitions may arrive at better judgements on policy issues and, hopefully, at 'better' policies and ways of delivering those policies.

Many commentators argue that the rational-analytic and argumentative-responsive approaches to policy evaluation seem to belong to different worlds (cf. Patton, 2002; Pawson and Tilley, 1997). Responsiveness to societal change and other actors' perspectives, however, is crucial for any intelligent results-oriented budgeting system. Applied in a system of results-oriented budgeting, where policy objectives and performance indicators play such important roles, the added value of policy evaluation should principally be the *systematization* of the possibility that dialogue and discussion will reveal differences (Majone, 1989). Building upon the main questions from the results-oriented budgeting system (Have we achieved what we intended? Have we done what we should have done in achieving it? Did it cost what we expected?) on which – in principle – information should be available on a regular basis, the main set of questions should be:

- Are we (still) on the right track?
- Are our assumptions (still) valid?
- Does the set of policy objectives and performance indicators focused on (still) represent what we must try to achieve?
- Which negative side effects of the use of policy objectives and performance indicators can be observed (e.g. are there indications that new, relevant information has been ignored or that strategic use of indicators and targets has frustrated performance)?

For the critical evaluator, the challenge is to facilitate the combination of analytical focus with the ability to continuously and critically review assumptions and performance indicators. Through interim 'verifying evaluations' that explicitly raise the validity of existing policy theories, targets and performance indicators yet capitalize on the advantage of focus can keep the risks manageable. Seeking information and insights from stakeholders (e.g. implementation offices, recipients), the evaluator may provide feedback and new information on the validity of policy objectives and performance indicators (cf. European Commission, 2004).

Experiences and New Evaluation Ambitions in the Netherlands

Experiences with results-oriented budgeting and the integrated use of evaluation research in the Netherlands have been cautiously positive. For a growing number of policy objectives, some form of effect and performance information is now included in budgets and annual accounts. Increasingly, policy objectives and performance indicators play important roles in the day-to-day management of

programmes. Ministers are being asked more frequently by the Lower House about the rationality behind policy proposals: the presumed causal relationship between the resources used, the instruments employed and the effects attained. Also, evaluation is used increasingly. Commitment to the results-oriented budgeting system is still strong: '(We must) continue working on Policy Budgets and Policy Accountability because we are definitely not there yet: there is still a great need for accessible budgets, for better accounting and greater efficiency' (Ministry of Finance, 2004: 53).

True, there has also been criticism. Many have pointed to the dangers of the political debate becoming 'bureaucratic', 'narrowed down to measuring' and 'vulnerable to strategic behaviour'. In addition, government (in the official response to the 2004 evaluation), Court of Auditors and Parliament all stressed the need to downsize the number of objectives and performance indicators in budgets and annual accounts.

Most importantly, despite indicators and mandatory policy evaluations, there is still inadequate insight into the effectiveness of a large portion of government expenditure. Evaluations tend to remain within the fixed boundaries of policy ambitions and theories: questions about the effectiveness or validity of existing policy programmes are rarely asked.

(T)he assumptions behind the achievement of objectives and the success of policy (for example, economic developments) are hardly discussed. It is often not clear why, and when, certain instruments or measures will help to achieve targets, and to what extent this will genuinely contribute to the objective. (Ministry of Finance, 2004)

As a consequence, ineffective policy programmes are allowed to continue, diverting funds and resources from potentially more effective ones.

To illustrate the need for verifying evaluation, some examples are presented here. In the past five years, several of the policy programmes included in the Netherlands state budget were subject to changes to their policy objectives and performance indicators. Table 2 summarizes the major changes.

In all these examples, the perceived lack of success forced policy-makers to rethink existing policy practices. In these processes, monitoring the output of government agencies and their results played an important evaluative role. In the example of 'biological agriculture', for instance, the lack of public demand for organic fruit and vegetables proved to be a formidable obstacle for farmers to change their businesses and, hence, convert acres from traditional production to biological production. In the domain of water management, several near floods in the 1990s led to the insight that providing rivers with extra space for overflow is an indispensable supplement to dykes in order to prevent disaster. In the field of education, continuing difficulties for members of ethnic minorities in finding employment meant the end of programmes promoting modules in ethnic minority languages and the beginning of special refresher classes in the Dutch language.

The main driver behind the policy changes in the domains featured in Table 2 is, as said, the perceived lack of success of existing policy programmes. Notably, the insight that roundabouts with separate cycle lanes (on which cyclists did not

Table 2. Policy-Oriented Learning: Changes in Paradigms, Objectives and Indicators

Policy area	Old objectives	Old indicators	New objectives	New indicators
1. Integration of ethnic minorities	Mastering a level of the preferred 'ethnic minority language'	Pupils attending after-school classes and successfully taking exams (nos. and %)	Mastering a set level of the Dutch language	Hours within regular teaching programme and 'end-level' in mastering Dutch (hours, exam results in nos. and %)
2. Biological agriculture	Making the supply-side of agricultural products more biological Keeping Holland dry	Acres of biological fields (nos. and %) Heights of dykes (metres, linked to high-water levels and likelihood of flooding)	Increasing the demand for biological products	Market share of biological products (%)
3. Water management	Keeping Holland dry	Heights of dykes (metres, linked to high-water levels and likelihood of flooding)	Keeping populated areas and land in commercial use dry	[Complementary to old indicators.] Space for controlled flooding (m ³ linked to river drainage capacity)
4. Transportation: traffic jams	Combating traffic jams	Traffic jams (nos., duration, measured in 'vehicle hours loss')	Creating reliability in transportation duration	Deviations from the average expected time (norm) (nos., deviation in minutes and hours, % of trips)
5. Green electricity	Increasing the demand for and use of green electricity	Market share of green electricity used for 2005(%)	Organizing an environmentally sustainable system of electricity supply and demand	Electricity that is produced in an environmentally sustainable manner (%)
6. Soil treatment	Cleansing of all contaminated soil	Contaminated spots cleansed (nos. and %)	Matching soil use and soil quality	Contaminated spots where use and quality are matched (nos. and %)
7. Road safety	Creation of roundabouts	Hot spots converted into roundabouts (nos. and %)	Equipping roundabouts with separate non-priority cycle lanes	Roundabouts equipped according to new insights (nos. and %)

enjoy priority) led to a spectacular 86 percent reduction in serious accidents was revealed in an evaluation! Unlike most of the other examples, where a persistent lack of success leads (or may lead) to a paradigm shift, this example illustrates the dynamic purpose of evaluation research: to help government (or society at large) avoid the negative consequences of sticking to the wrong policy measures for too long. I strongly believe that we must use policy evaluation in this way to facilitate (the speed of) the learning processes that can lead to necessary paradigm shifts. As was made clear in the 2004 *VBTB* evaluation, we cannot take it for granted that evaluation automatically takes up this role. And although, in the end, it will always remain the responsibility of civil servants and politicians to ask the fundamental questions, planning responsive, verifying evaluations will increase the likelihood that they are being asked.

From this perspective, at the request by the Netherlands National Spatial Planning Agency (NSPA), Teisman and Van der Meer have formulated several practical rules to select the most appropriate 'evaluation arrangement' for a specific policy phase (NSPA, 2002). The objective is to deliberately create uncertainty about the chosen direction and performance indicators. As well as careful organization and timing, Teisman and Van der Meer call for policy evaluation that will do the following:

- Distinguish between and interlink static and dynamic policy evaluation: 'Accountability requires an assessment of what has become of the original objectives and agreements. It should also respond to new, emerging needs and insights.'
- Combine quantitative and qualitative policy evaluation: 'Depending on the subject, qualitative considerations are necessary (for example on spatial quality) or, alternatively, quantitative analyses based on tightly formulated indicators.'
- Consciously deal with product and process evaluations: 'Product evaluation asks what effects the policy has had. Process evaluation asks how the policy and implementation processes were conducted. Both forms are necessary.'
- Enable stakeholders to systematically reflect on policy opinions, in part by comparing qualitative and quantitative evaluations with each other and by combining internal and external evaluations: 'An evaluation becomes more significant if it is based not only on policy effects but also on insights into how they were (or were not) achieved' (NSPA, 2002: 9–12; van der Meer and Edelenbos, 2006).

The NSPA's strategy signifies a development into a dynamic use of objectives and indicators. More importantly, the need to gear policy evaluations to the desired level of learning is also echoed in the Ministry of Finance's new plans for 'policy analysis' (Ministry of Finance, 2004). The following questions will be asked in respect of a particular field of policy (defined as the policy aimed at a general or operational objective):

- What is the problem behind the policy? Is this problem still current? What is the cause of the problem?

- What objective has the government formulated for the solution to the problem?
- What instruments are being used? (To what degree?) Do the instruments contribute to the objectives formulated (to a solution of the problem)? What are important positive and negative side effects?

By asking these questions in a systematic way, civil servants and politicians alike will have ample opportunity to critically reflect on existing policy objectives and performance indicators. This is especially relevant when dealing with complex program logic of social dynamics (de Lancer Jules, 2006). The new proposed strategy, however, does lack two important elements. First, there is nothing in the proposals on the consultation of stakeholders. The Ministry of Finance's ambition to work with expert committees indicates that this omission will be repaired in practice. Second, there is no explicit reference to the negative impacts of policy objectives and performance indicators. Here, too, the ongoing debate on the merits and risks of indicators and targets warrants optimism (e.g. the critical report of the Scientific Council for Government Policy, Scientific Council, 2005).

Conclusion

Policy evaluations should contribute to the quality of public policies. In a democratic society, there is an additional link: policy evaluations should also contribute to the quality of policy-oriented dialogue and deliberation. In part, this can be done by enabling the constructive yet discerning use of performance indicators and policy objectives. I firmly believe that casting ambitions in the form of policy objectives and performance indicators is essential to a learning and accountable government that is willing to engage in a substantive policy dialogue.

Responsive, verifying evaluation constitutes a critical instrument in overcoming the downsides of policy objectives and performance indicators. One of its key aims should be to maintain discussion of the original indicators and targets *without* compromising their value and significance. As well as adding *ex ante* to the quality of policy formulation and *ex post* to learning processes on behalf of policy-makers, policy evaluation has a dynamic function. By actively engaging in these discussions and seeking out stakeholders' opinions and motives, the evaluator helps government to remain sufficiently responsive to external changes and developments while retaining the locus, focus and language provided by policy objectives and performance indicators.

Note

1. See Sterck et al. (2005) for an extensive overview of the evolutions in the control pyramid and performance budgeting of six OECD countries (Australia, Canada, the Netherlands, Sweden, the United Kingdom and the United States).

References

- Abma, T.A. (1996) *Responsief evalueren: Discoursen, controversen en allianties in het postmoderne*. Delft: Eburon.
- Bock, D.E. de, R.J. in 't Veld, R.C.D. Berndsen et al. (1996) *Beleidsvaluatie: Zwart gat en dynamiek*. The Hague: Reed Business Information.
- Bovens, Mark and Paul 't Hart (2004) 'Publieke verantwoording: Zegen en vloek', in W. Bakker and K. Yesilkagit (eds), *Publieke verantwoording*, Amsterdam: Boom, 2005, pp. 245–64.
- Bruijn, Hans de (2001) *Prestatiemeting in de publieke sector*. Utrecht: Lemma.
- Court of Auditors (2005) *Staat van de Beleidsinformatie 2005 [State of Policy Information 2005]*, The Hague: SDU.
- de Lancer Jules, Patricia (2006) 'Performance Measurement: An Effective Tool for Government Accountability? The Debate Goes On', *Evaluation* 12(2): 219–35.
- Dryzek, J.S. (1982) 'Policy Analysis as a Hermeneutic Activity', *Policy Sciences* 14: 309–29.
- Dunn, William N. (1994) *Public Policy Analysis*. Englewood Cliffs, NJ: Prentice Hall.
- European Commission (2004) *Guide to Evaluating Socio Economic Development*. Brussels: EC.
- Everitt, Angela (1996) 'Developing Critical Evaluation', *Evaluation* 2(2): 173–88.
- Fisher, Frank and John Forester (1993) *The Argumentative Turn in Policy Analysis and Planning*. London: UCL Press.
- Gerritsen, E. and L. Geut (1997) 'Beleidsvaluatie: Zwart gat en dynamiek', in *Naar een doelmatiger overhead: Platform Doelmatigheid*. The Hague: Nivra, pp. 34–47.
- Guba, Egon G. and Yvonna S. Lincoln (1989) *Fourth Generation Evaluation*. Newbury Park, CA: SAGE.
- Hood, C. (2002) 'The Risk Game and the Blame Game', in *Government and Opposition*, 2002/1, pp. 15–37.
- In 't Veld, R.J. (1989) *De verguisde staat*. The Hague: VUGA.
- Kraan, D.J. (2003) 'De institutionele economie van de verzelfstandiging', in A.P. Ros and H.R.J. Vollebergh (eds) *Ontwikkeling en overhead*. The Hague: SDU.
- Majone, G. (1989) *Evidence, Argument, and Persuasion in the Policy Process*. New Haven, CT: Yale University Press.
- Ministry of Finance (1991) *Subsidies: Een analyse van de actualiteit, effectiviteit en doelmatigheid*. The Hague: SDU.
- Ministry of Finance (1999) *Van beleidsbegroting tot beleidsverantwoording [From Policy Budget to Policy Accountability]*, The Hague: SDU.
- Ministry of Finance (2001) *State Regulation on Performance Data and National Government Evaluation Research*. The Hague: Ministry of Finance.
- Ministry of Finance (2004) *Eindrapport VBTB Evaluatie: Lessen uit de praktijk*. The Hague: Ministry of Finance, Dec.
- National Spatial Planning Agency (NSPA) (2002) *Evalueren om te leren: Naar een evaluatiearrangement voor het ruimtelijk beleid*. The Hague: Ministry of Housing, Spatial Planning and the Environment/RPD (written by G.R. Teisman and F.B. van der Meer; summary by Jelle Troelstra).
- OECD (2001) *Government for the Future*. PUMA Policy Brief, 9, Paris: OECD, June.
- OECD (2004) *Public Sector Modernisation: Governing for Performance*. OECD Observer, October.
- Patton, M.Q. (2002) *Qualitative Research and Evaluation Methods*. Thousand Oaks, CA: SAGE.

- Pawson, R. and N. Tilley (1997) *Realist Evaluation*. London: SAGE.
- Perrin, Burt (2006) *Moving from Outputs to Outcomes: Practical Advice from Governments around the World*. Publication based upon the Worldbank Roundtable 'Moving from outputs to outcomes'. Washington, DC: IBM Center for the Business of Government.
- Pirsig, Robert M. (1992) *Lila: An Inquiry into Morals*. New York: Bantam Press.
- Power, Michael (1997) *The Audit Society*. Oxford: Oxford University Press.
- Rein, Martin, and Donald A. Schön (1993) 'Reframing Policy Discourse', in Frank Fisher and John Forester (eds) *The Argumentative Turn in Policy Analysis and Planning*. London: UCL Press.
- Schwandt, T.A. (2001) *Evaluation Practice Reconsidered*. New York: Peter Lang.
- Schwandt, T.A. (2002) 'Democratization, Globalization, and Performativity: Implications for Learning in and from Evaluation', keynote address given at the fifth biennial meeting of the European Evaluation Society, Seville, 12 Oct. (<http://www.europeanevaluation.org>).
- Schwartz, Robert and John Main (2005) *Quality Matters: Seeking Confidence in Evaluating, Auditing, and Performance Reporting*. New Brunswick, NJ: Transaction Publishers.
- Scientific Council for Central Government Policy (Wetenschappelijke Raad voor het Regeringsbeleid) (2005) *Proofs of Good Service Provisions*. The Hague: WRR.
- Sorber, Bram (1999) 'Performance Measurement in the Central Government Departments of the Netherlands', in Arie Alachmi (ed.) *Performance and Quality Measurement in Government*, Burke, VA: Chatelaine Press.
- Stake, R.E. (2004) *Standards-Based and Responsive Evaluation*. Thousand Oaks, CA: SAGE.
- Sterck, Katrien, Bram Scheers and Geert Bouckaert (2005) *The Modernisation of Public Control Pyramid: International Trends*. Leuven: Public Management Institute, Catholic University of Leuven.
- Van der Knaap, Peter (1995) 'Policy Evaluation and Learning: Feedback, Enlightenment or Argumentation', *Evaluation* 1(2): 193–220.
- Van der Knaap, Peter (2000) 'Performance Management and Policy Evaluation in the Netherlands: Towards an Integrated Approach', *Evaluation* 6(3): 335–50.
- Van der Knaap, Peter (2004) 'Theory-Based Evaluation and Learning: Possibilities and Challenges', *Evaluation* 10(1): 16–34.
- van der Meer, Frans-Burke and Edelenbos, Jurian (2006) 'Evaluation in Multi-Actor Policy Processes: Accountability, Learning and Co-operation', *Evaluation* 12(2): 201–8.
- Weick, Karl E. (1979) *The Social Psychology of Organizing*, Reading, MA: Addison-Wesley.
- Wildavsky, Aaron (1987) *Speaking Truth to Power: The Art and Craft of Policy Analysis*. New Brunswick, NJ: Transaction Books.
- Wit, B. de, R. Meyer and K. Breed (2000) *Strategisch management van publieke organisaties: De overheid in paradoxen*. Utrecht: Lemma.

PETER VAN DER KNAAP is director of Policy Evaluation Research at the Netherlands Court of Auditors. Prior to this he held positions at the Netherlands Ministry for Finance and Erasmus University Rotterdam. He helped launch the result-oriented budgeting system in the Netherlands. His main research interests include theory-based policy evaluation, constructivism and organisational learning. Please address correspondence to: Netherlands Court of Audit, Lange Voorhout 8, 2514 ED, The Hague, The Netherlands. [peter.vanderknaap@rekenkamer.nl; peter.vanderknaap@wanadoo.nl]