

Evaluation in multi-actor policy processes: accountability, learning and cooperation

Frans-Bauke van der Meer

Jurian Edelenbos

ABSTRACT

Two main functions of evaluation are to enable accountability and collective learning. Both of these – and their combination – run into divers complications when applied in complex multi-actor policy processes. The article explores these complications and illustrates these with examples from the field of spatial policy. In doing so a third function of evaluation in such contexts is identified, viz. evaluation as an instrument of cooperation. Next, a number of theoretical ideas, supported by empirical research, are proposed in order to understand better when, why and how evaluation contribute to complex multi-actor policy processes. Based on these insights some principles are elaborated for the development of constructive evaluation arrangements. It is suggested that cooperation is a precondition for preservation of accountability and learning functions of evaluation in multi-actor settings.

Key words: evaluation, accountability, learning, cooperation, spatial policy

Correspondence address

Dr. F.B. van der Meer, Center for Public Management, Faculty of Social Sciences,
Erasmus University, PO Box 1738, NL-3000 DR Rotterdam, The Netherlands.

Phone +31 10 408 2137; E-mail: f.b.vandermeer@fsw.eur.nl

Evaluation in multi-actor policy processes: accountability, learning and cooperation

Introduction

In general evaluation can be ascribed two main functions in policy processes. First, it supports and facilitates accountability and transparency by assessing policy output and outcomes (and comparing these with policy goals), by assessing the efficiency of policy implementation and sometimes also by assessing the extent to which actual results can be ascribed to the policy. Second, evaluation may contribute to learning processes leading to improved policy making and/or implementation. Here the same features are relevant as well as the assessment of how and why policy relevant results come about; such theory informed explanations may provide new cues for policy development and improvement. Still, there may be contradictions between the two functions mentioned. For example, if the emphasis is (perceived to be) on accountability, evaluation may invoke defensive reaction, which may hem learning processes.

Now, policy-making and implementation increasingly is shaped by the interplay between efforts (concerted or not) of different actors, both from public and private spheres. Although this is not altogether new in many policy domains, the extent to which both policy-making and implementation are explicitly conceived of in terms of cooperation is certainly growing. The question how evaluation ‘works’ and how it

can fulfill and combine its functions in multi-actor policy processes is therefore highly relevant.

In this contribution we address this question by proceeding as follows. First, we discuss a number of shifts in the nature of policy processes and identify a number of key issues related to the functions of evaluation in multi-actor policy processes. Then we present an exemplary case, viz. Dutch spatial planning (1) and explore the actual and potential functioning of evaluation in this field. Finally, on the basis of our analysis and experiences with new forms of evaluation, we suggest some directions in which both individual evaluations and combinations of related evaluations may be developed and organized in order to improve and combine their different functions in multi-actor contexts. Our findings and suggestions are summarized in a number of principles, which may be useful for other multi-actor domains as well. We will suggest that different forms and measures of cooperation may be vital to preserve and combine the functions of evaluation in such policy processes.

Setting the problem: how to meet contradicting demands in evaluation?

Accountability

Looking at the role of evaluation in present policy processes, one may get the impression that we witness a contradictory development. On the one hand there is a still increasing emphasis on transparency, measurable results and accountability. Policy documents should specify clear goals, the attainment of which should be measured by unequivocal (and if possible quantitative) indicators. Policy makers should be hold

accountable for the results thus assessed. Evaluation, therefore, should assess efficiency, output and outcomes of policies against their (initial) goals (Provus 1971; Stufflebeam 2001).

On the other hand policy making as well as implementation is increasingly seen in terms of complex interactive processes involving many actors. Moreover, it is generally accepted that the dynamics of societal problems is only partly influenced by public policy. Other social-economic, technological, cultural and societal developments play a role as well. Moreover, nowadays there often seems to be much ambiguity in and around policy processes: lack of shared goals, lack of information (or abundance of information but lack of meaning) on what the problems are, what causes them, how possible measures will work out, et cetera (Scharpf 1997).

For a number of reasons these latter features of current policy processes seem to generate substantial obstacles for the accountability ambitions. First, the multi-actor and interactive nature of many policy processes implies that it may not be self-evident who is to be held responsible for what (Mayne 2003). If policies are the result of negotiation and collaboration between different stakeholders, then who is responsible for eventual results? What if a minister decides on general policy lines, but the elaboration is commissioned to an autonomized agency? And what if policy implementation is partly done by other public or private bodies than those involved in policy making, and still others are charged with supervision and control? Sometimes one tries to solve this problem by defining distinct responsibilities for each actor involved. But where a policy result aimed at can only be reached by actual cooperation it may be hard to define and ascribe separate responsibilities (Klijn & Teisman 2001). Under such circumstances a result can only be realized by cooperation and shared responsibility. This

requires a pluricentric evaluation approach, in which people recognize the notion that objectives are the outcomes of an interaction and argumentation process between various interested parties (Guba and Lincoln 1989; Majone 1989; Fischer and Forester 1993; Dryzek 1993).

Second, through the complexity, ambiguity and interconnectedness of policy and societal processes it may be problematic to attribute societal developments to related policies. A positive 'result' may well have different causes, while a negative result might have been even worse if the policy had not been applied.

Third, both the multi-actor nature and the ambiguity of many policy processes imply that it will often not be possible to define clear criteria which relate in a sensible way to the quality of the policy and on which consensus can be reached. Moreover, such policy processes will often be more of a goal seeking than of a goal driven nature; goals change as processes proceed and therefore also the focus of the evaluation evolves. This shift in focus is a result of gaining new insight on the way, and is therefore a result of learning (Guba and Lincoln 1989; Van der Knaap 1995). We explore this function of evaluation more closely in the next section.

Learning

The second function of evaluation, learning, seems to be more important in complex multi-actor situations compared to more simple and/or more hierarchical environments, simply because there is more uncertainty and ambiguity around policies and hence more to be learned about their impacts and dynamics (Scharpf 1997).

Also a different kind of knowledge is needed here. Of course, feedback on the efficacy of policy instruments, the efficiency of its administration, and the validity of the

presuppositions behind the policy remain useful and necessary if policy-making and implementation is to be improved. But in the multi-actor context there are many actors with ‘policies’ and ‘instruments’, the validity, efficiency and efficacy of which is dependent on their mutual interplay (Kickert et al 1997).

Moreover, these actor behaviors are – at least to a considerable extent – determined by their interpretations of the situation, of actions of other actors and of perceived results of previous actions. Effective policy making and implementation, then, requires also insight in patterns of interpretations actors make and in the dynamics of their interactions (Van der Meer 1999).

For evaluations that are to support learning in multi-actor contexts, this may have two implications. One is that in its explanation of policy results these processes of sensemaking and interaction should be taken into account. This may conflict with accountability, because such an evaluating analysis may show that the policy actually is the result of co-production which cannot be controlled in a strict sense by one focal actor held accountable. The other implication is that the analysis in an evaluation should make sense seen from the perspective of actors involved, if they are to learn from it. Actor interpretations and interactions do not only play a role in the production of (outcomes of) policies, but also in the generation of learning processes as a consequence of evaluation (Majone 1989; Fischer and Forester 1993).

Cooperation

A further complication that can be related to the functions of evaluation in multi-actor processes is that learning by some or all actors involved does not necessarily lead to improved policies. There is a collective action problem. One can imagine that differ-

ent actors learn to act more effectively, each from his own perspective, and decide to act opportunistically in trying to realize their own interests. The net result may be that improved strategies neutralize each other or produce undesirable effect by their mutual interplay. So cooperation and coordination is needed.

We argued before that policy-making is a process of collaboration. Evaluation should not be treated otherwise. Evaluation is also a product of cooperation. In literature on evaluation we see a development towards participative, constructivist, and responsive types of evaluation (Stake 1983; Schwandt 1984; House and Howe, 2000; Guba and Lincoln 1989). To various degrees, evaluation studies were conducted together with those involved.

But more important, evaluation is not the end station of a collaborative policy making process, it is input for continued and deepened cooperation. Evaluation provides learning aspects and facilitates cooperation between stakeholders. The challenge is then how to consolidate cooperation, which was realized in the constructive evaluation process.

We discussed three possible functions of evaluation, supporting accountability, facilitating learning and helping cooperation between stakeholders. These functions put different and often contradicting demands on evaluation. The following questions for this article pops up: how can the contradicting demands on evaluation be reconciled? How can evaluation help to identify accountable actors and their domains of responsibility, and at the same time facilitate learning by different actors (with their diverse perspectives) and realize collective action?

In order to explore the main question formulated in this section, we now turn to the case of Dutch spatial planning in order to illustrate this question. The policy processes involved will be introduced in the next section. After that, we will identify problems with respect to evaluation in the specific domain of spatial planning and then turn to possible directions for their solution.

Dutch spatial planning

The way in which the Dutch spatial planning policy gets shape and is implemented matches the situation depicted in the previous section. We will give a short description of patterns and traditions in spatial planning practice in the Netherlands and of the steering model behind the recent policy statements of the Dutch central government on spatial planning, the so called Fifth Bill (VROM 2001) and the Bill on Space (VROM 2004).

In the spatial domain the decentralized unitary character of the Dutch government system reflects itself clearly. A decentralized unitary state means in principle that we have a hierarchic steering system: the central government. But at the same time, the two other levels, regional and local government, have their own responsibilities and tasks. Each level has tasks and competencies that are considered appropriate to be addressed at that level. These tasks are carried out with a certain amount of autonomy, but at the same time have to be carried out by certain rules formulated and monitored by the central government (Woltjer 2000).

With respect to spatial policy, central government develops policy on infrastructure, housing and regional development, but structure and zoning plans are made locally by the municipalities and supervised regionally by the provinces. Rules, conditions and criteria are formulated on each of the levels of government. Of course, decentral developments have also consequences on higher levels. Regularly this gives rise to calls for more centralization. The result has been in part disappointing, however, since central steering appeared to be limited in its efficacy.

A further complication in the spatial domain is that at all these levels also private actors (enterprises, interest groups) participate one way or another in the policy-making process. Increasingly they are invited to do so by governments through organizing public private partnerships (Klijn and Teisman 2001) and interactive policy-making processes (Edelenbos 1999). Public private partnership is cooperation between public and private actors with a durable character in which actors develop mutual products and/or services and in which risk, costs, and benefits are shared. Interactive policy-making is cooperation between governments and societal groups and citizens. The Fifth Bill has been subject of a broad interactive process, called “Give Them Space”. Citizens have been consulted through surveys. Moreover, regional workshops have been organized in order to give opportunity to give input for the creation of the Bill. The interactive process was finished by a closing conference in June 2000. Next to these kind of interactive processes, there are many opportunities and provisions for participation or legal objection in Dutch legislation, especially in the spatial planning domain. Individual citizens and interest groups often intensively use these opportunities.

Private actors play an important role in spatial development. Private actors initiate an important part of the building projects. They also finance the major part of it. The physical realization is almost entirely done by private actors. With respect to infrastructure, the situation is somewhat different, but here too private actors do part of the financing and nearly all of the realization.

The total need for space is enormous in The Netherlands. The Fifth Bill on Spatial Planning estimates that The Netherlands should in fact be 10-15% bigger than it actually is in order to accommodate the aggregated claims. With approximately 470 inhabitants per square kilometer in 2001 the Netherlands is also one of the most densely populated countries in the world. Land is a scarce commodity, especially in the western, most urbanized part. This part of the country, called the Randstad or Ring City, accounts for approximately 6,5 million inhabitants and is rapidly developing into one metropolitan area. As a consequence, especially in this part of the country problems of congestion and rocketing prices for land and real estate become dominant. Infrastructure, industry, housing, agriculture, nature, and recreational spaces have to be located carefully and attuned to one another. Spatial planning is vital. National, provincial and municipal governments each have their own powers for spatial planning. They are also dependent on each other. Agreements need to be reached between the various levels of government to ensure that plans harmonize. Moreover, people have many - often contradictory - wishes with respect to the quality of the environment with respect to nature, to recreation, to living, to rest, to fast transport, etc.

And even if government is able to define a ‘general interest’ it is often not exactly clear what it should do or what the impact of certain measures will be, because there are all kinds of intricate relationships and contingencies that have to be taken into account. Traffic congestion may be related to the perceived quality of public transport. But what will happen if the tariffs are made lower or the speed of trains is enhanced? And how will more or broadened highways work out? A government can decide to develop a new living area or to promote a new type of houses in which living and working can be combined, but it is dependent on the preparedness of citizens to go and live there. Actually, in the framework of the Fourth Bill on Spatial Planning (VROM 1988; VROM 1992) central government defined a number of such new areas, which was not altogether a success (Werkgroep Vijfde Nota Ruimtelijke Ordening 1999-2000).

So, when a new bill on spatial planning was to be drafted, there was a dilemma. On the one hand it was felt that effective measures had to be taken in order to accommodate the conflicting spatial claims, to prevent further messiness in both urban and rural area, and to enhance spatial quality. On the other hand there was the feeling that directive central policies might not work. From experiences in the framework of the Fourth Bill many have concluded that local and regional authorities should not be merely considered implementing agencies, but should get more opportunities for own spatial policies, adapted to local circumstances and needs of local actors.

Eventually, the Fifth Bill announced a new steering model, which can be summarized as follows:

- the bill is a strategic policy statement, a general vision on main points and desirable future developments. Protection of nature and urban development are the cen-

tral focus points, which together require a combination of control and development strategies;

- the specification and implementation of spatial policy is not a hierarchical process: regional and local governments actually should make policy;
- local and regional authorities are expected to orient themselves on the central policy lines and to coordinate their policies and plans among themselves. Provinces are seen as major coordinating and orchestrating bodies in this connection.
- central government only intends to ‘realize’ certain spatial projects of ‘national interest’. For the remainder it adopts a stimulating and facilitating role, amongst others by subsidizing local and regional projects. In doing so it tests local and regional plans against a number of guiding principles and general criteria for spatial quality (spatial diversity, economic and societal functionality, cultural diversity, social justice, sustainability, attractiveness, human scale);

So, the elements of the Fifth Bill on Spatial Planning are not only elements of the horizontal steering paradigm (bottom up processes), but also aspects of the vertical steering paradigm (top down processes). This blend has consequences for the existing institutional structures of accountability in the public sector. Accountability is traditionally seen as applying to hierarchical relationships with subordinates accountable to their superiors (vertical accountability). Collaboration, however, often involves a relationship built on non-hierarchical structures. So a more horizontal kind of accountability comes into play. How this will complement vertical accountability is not yet clear, but may become an additional or alternative form of accountability demanded by citizens. Collaboration among governments and citizen(s) (groups) pose

several accountability problems for governments. One of them involves the notion of "dual accountability" - *vertical accountability* to political institutions (like Parliament or city councils) and *horizontal accountability* to the collaborative entity itself (Mayne 2003).

Recently, the Bill on Space replaced the Fifth Bill in 2004. In this Bill even more room is given for decentralized public bodies, municipalities and provinces, for spatial development. However, some degree of centralization will remain. The tensions between horizontal and vertical accountability structures will therefore also remain an issue.

Evaluation in spatial planning policy processes

In our research (2) we identified a number of specific problems with respect to evaluation in Dutch spatial planning.

Accountability

In the framework of accountability processes, evaluation is complicated by the complexity, by the decentralization and by the multi-actor nature of spatial policy processes. The policy goals formulated at the central level are, as we saw, general and abstract. Their realization cannot be assessed unequivocally without operationalisation of these abstract goals. This operationalisation, however, is to a large extent done by decentralized governments. Still, central government can be hold accountable for national spatial quality. But which (or whose) operationalizations and criteria should an

evaluation use to make an assessment? One could try to reduce this problem by formulating more specific criteria at the central level. But that might be counterproductive. For example ‘spatial diversity’ at the national level is *reduced* if all local governments are forced to pursue the same diversity at the local level. To realize diversity at the national level it may be desirable to stimulate even more instead of less differentiation of goals between local areas, so that e.g. not every municipality is investing in its own large industrial area. This could be a central government task. But the interesting thing is that in present central policy the responsibility for adjustment of policy between adjacent territories is delegated to these territories – and for good reasons! But how, in view of all this, could evaluation be shaped to support accountability at the central level? It may be tempting here to switch the focus from ‘substantial’ results to process criteria. The Fifth Bill actually specifies procedural steps that should be taken and time schedules that should be met. For example, municipalities should define borders outside which in principle no building will be allowed (‘red contours’). They are required to have done so by the end of 2005. Policy success, then, can perhaps be measured by assessing whether such process criteria are realized: have the ‘red contours’ actually been drawn within the specified period? But still, there is a genuine need to be able to assess and to judge substantial policy results (3).

On the decentralized level there are problems too. Of course, if local policy goals are formulated clear enough the extent to which they are realized in a certain period of time, can be assessed. But here, like on the national level, the actual results are heavily dependent on what other actors do. We already mentioned the necessity to negotiate with adjacent municipalities and regional bodies. National government also influences local policies because the implementation of many projects is only possible

with subsidies by the national government. As we saw in the previous section, national government uses its own interpretation of national goals and local initiatives to decide on applications. Finally realization of spatial projects depends very much on the behavior of private enterprises and of individuals (both as citizens and as consumers). So the question to what extent local politicians can be held accountable for actual spatial developments, is a very intricate one.

A further complication is that – even if initial goals and responsibilities are clear – preferences of some or all actors may change over time. The same holds for economic conditions and technological developments. Since spatial projects may take many years, this is not a small problem. E.g. should infrastructure policies be judged by the extent to which projects once formulated are realized or by the extent to which the project (implementation) is adapted to changed circumstances? And if so, against which criteria?

Learning

There are many examples of spatial policies that fail, e.g. creation of industrial areas chosen by (too) few enterprises as their location; infrastructure depending on private financing which doesn't show up; new housing locations with much vacancies. There are also many spatial projects realized in a technical sense, but not providing the intended functions or considered a mistake after their realization (insecure housing areas, infrastructure creating environmental problems or dislocating in stead of solving congestion, underutilized infrastructure). And there are successes despite much initial opposition (carless areas in inner cities). The last example suggests that there may also be projects turned down although they might have been successful.

Taken together, these observations underline the necessity of learning in spatial policy about policy implementation and impacts, about private actor behavior, about production of unintended effects, about changes in patterns of values, preferences and objectives, etc. Given the enormous amounts of money involved, the irreversible nature of many spatial changes, and the societal weight of certain impacts, prevention of failure and more careful guidance of developments are necessary. Learning from past experience, and thus (ex post) evaluation, can contribute to the insights required. If evaluation is performed while policy-making processes proceed, a more 'learning by doing' and 'doing by learning' style is required.

However, specific learning needs will be different for different actors. This is not only so because they have different goals (e.g. the central housing department will be interested in effective ways of realizing house building projects, whereas municipalities want to know how to develop safe and attractive quarters). They also have (or perceive) different action domains and 'given conditions'. So, central government will be interested in aggregated data to discover trends and overall result, while municipalities will need disaggregated information on the production of specific impacts. Furthermore, each actor will primarily want to learn about the impacts of the specific instruments he has at his disposal. This has two implications for the relation between evaluation and learning. The first is that not all actors will perceive a certain evaluation as equally relevant and usable. This need not be a problem in itself, but it may complicate cooperation and mutual adaptation (see further below). The second implication is that certain insights from evaluation may not lead to learning effects at all if they do not to some extent fit in existing institutional patterns and in the distribution of competences. E.g. support for private initiatives may appear to be an effective way

of promoting safe and livable areas, but still not be practiced very much because of complicated rules and procedures or shortage of locally available funds. Hence, it is not at all self-evident that evaluations get effectively linked to actor perceptions of the situation, which would be necessary to produce learning effects (Van der Meer 1999).

A further limitation of learning effects of evaluation in the spatial domain is related to the interplay between its functions. Evaluations primarily having an accountability function tend to be public and are frequently performed by external evaluators. Units whose policies, management or implementation activities are being evaluated will often be inclined to defend what they have done and achieved. This orientation reduces chances for learning from the evaluation about how things can be done better. This is a fortiori the case in complex multi-actor contexts, since here many actors will see good reasons to discredit a specific evaluation. If for example a review of spatial developments at an aggregated nation wide level is confronted with goals in the central government policy bills, provinces and municipalities probably won't feel themselves responsible, although they in fact contribute to a very large extent to these developments. This is so because they do in part have different goals, experience different boundary conditions and interdependencies and act on a different level of aggregation. Evaluations more valid from the point of view of decentralized governments may in turn be thought of little value at the central level. This stresses the need for more participatory forms of evaluation (House & Howe, 1998; 2000).

Evaluations aimed at learning in a multi-actor context may bear the perceived risk to be used for accountability. E.g. higher levels may pick up local assessments made to support a debate on possible new policies as a basis for judgment of local performance. In anticipation of such 'misuse' actors may be inclined to perform inter-

nal, not published evaluations if they want to learn from it. These conditions may enhance learning on the one hand (because actors will be more open for evaluation results and their implications) but limit learning on the other hand (because the external critical view is missing and little 'new' things will be found).

Cooperation

The dominant accountability trend and the differences of goals and perspectives between spatial policy actors gives rise to a large diversity of evaluations, not only at different administrative levels but also initiated by private or societal organizations, such as the environmental movements. All kinds of (seemingly) contradictory results are to be expected, which may block not only (collective) learning, but also cooperation itself.

So, certainly in the field of spatial development, our question how the different evaluation functions can be realized and combined in complex multi-actor policy processes appears to be a very urgent one, not only to keep evaluation alive, but also enable successful policies as such.

To find some starting point for answers to this question, we first make one step back and ask: what is it that makes evaluation influence policy processes?

Utilization and impact of evaluation

At first sight evaluation is a pervasive phenomenon in public administration. Evaluation studies in a broad sense, including audits and other assessments, are performed frequently and typically laid down in reports to be used in consecutive planning and decision-making. Subsidies granted by government often are accompanied by the request that an evaluation should be done after a fixed period of time. Also arrangements with respect to evaluation frequently are inserted in political compromises and package deals. Incidents and calamities often give rise to (many) evaluation and as-

assessment studies. Moreover institutions like the Court of Audit or universities may initiate all kinds of evaluations on their own account. Many times, their reports catch extensive attention in press or parliament.

Still, there is much uncertainty and uneasiness about relevance and impact of many evaluation efforts. Often evaluation reports are perceived to disappear in proverbial bureau drawers (Patton 1997, p. 7-10). Weiss (1980) gave a far more subtle account of the apparent absence of evaluation impacts. On the basis of interviews with decision makers she argued that evaluations often do not have direct traceable impacts but do nevertheless influence the state of mind of decision makers, which in turn shape their small decisions (that often go unnoticed, perhaps even by the decision makers themselves). Thus she suggests a ‘creeping’ learning process. But even if she is right, the mechanisms that produce or prevent evaluation impacts on such learning processes have to be investigated further, if we are to understand eventual impacts and to develop more effective evaluation practices.

One question is *whether, when* and *by whom* evaluations are used. This question has given rise to an extensive ‘utilization’ literature (Shulha and Cousins 1997, p. 196).

The cited authors observe that this research

“... produced an extended “shopping list” of likely predictors [of utilization].

These predictors tended to cluster into categories such as: (a) relevance, (b) credibility, (c) user involvement, (d) communication effectiveness, (e) potential for information processing, (f) clients’ need for information, (g) anticipated degree of program change, (h) perceived value of evaluation as a man-

agement tool, (i) quality of evaluation implementation, and (j) contextual characteristics of the decision or policy setting" (Shulha and Cousins 1997, p. 196).

The last category may refer to factors such as the timing of reports in relation to the actual stage of a policy process (Mulder et al. 1991; Rist 1994, p. 194-199) and whether or not there is an institutionalized evaluation practice (Leeuw and Rozendal 1994).

Efforts to substantiate the contribution of any or all of such factors or conditions are frequently not very conclusive or even contradictory (cf. also Leeuw and Rozendal 1994; Shulha and Cousins 1997, p. 196-197). The general impression is that the factors identified will be helpful to enhance utilization, but that they are neither necessary nor sufficient.

In part this finding may be ascribed to the fact that perhaps not so much 'objective' characteristics (e.g. quality or credibility) but actor's perceptions of those characteristics matter (see e.g. Rist 1994, p. 200-203; see also Child 1972, p. 4-5). This gives rise to the question how perceptions come into being. In this connection increasing attention is given to the role that is played by contextual factors and by the nature of the relations between evaluator and stakeholders (Shulha en Cousins 1997). Indeed, the idea that stakeholders should recognize their ideas and experiences in evaluation studies if they are to do something with the results, is at the core of many new modes of evaluation practice, such as 'utilization-focused' evaluation (Patton 1997), 'participative' evaluation (Cousins and Earl 1992) or 'responsive' evaluation (Stake, 1983; Abma 1996) and 'fourth generation' evaluation (Guba and Lincoln

1989). Thus, in the development and application of evaluation methods and practices, an increasing focus on ‘the user’ can be observed. Do such practices and methods enhance the utilization of evaluation results? And, if so, how come? What are the mechanisms in these and other forms of evaluation that influences perceptions and behavior of actors in the direction of more (or less) utilization?

Perhaps even more important than the question when evaluations are used is the question *how* and *by whom* they are used. When and how do evaluations contribute to which changes in policies, policy-making, strategies, implementation or organization in the public domain? And when and how are evaluations ignored or mainly used to strengthen existing positions? Here the focus should be on the *mechanisms* that produce different types of impact.

In order to find answers to these questions we have to make an interpretation of the nature of our problem and devise a theoretical and methodical approach fit to tackle it. We make a number of presuppositions.

First, we presuppose that human behavior is related to ‘sensemaking’ (Weick 1995). That is to say that the ‘meaning’ (Mead 1962 (1934)) situations, incidents or evaluation reports have, or get, will be constitutive for (in)action of the actor. In other words we start from the premise that it is not so much an evaluation ‘as such’ that evokes effects, but the way it is perceived and interpreted by actors in the field. This does not imply that sensemaking always (or only) precedes action. On the contrary, we think that (experiences in) action, however produced, shape (but not determine!) meaning (see Weick 1979, p. 194-201 on ‘retrospective sensemaking’).

Second, we presuppose that the production of both ‘meaning’ and ‘action’ are patterned and socially contextualized. That is, sensemaking and emerging behavioral patterns reflect (sub)cultures, previously institutionalized styles and values, which, in turn, are historical products of preceding interaction, sensemaking and behavior (Selznick 1984 (1957); March and Olsen 1989; DiMaggio and Powell 1991). Hence, although we view institutions as social constructions, embedded in worldviews of actors, they will often and to a large extent have a given and ‘out there’ character for actors, just because they are embedded in their worldviews. However, the way we conceive of the generation and continuation of institutions implies that they may change, often gradually, in and by ongoing interaction and experience.

Third, we start from the premise that the contexts in which evaluation studies and their impacts are produced are of a multi-actor nature. The contexts can, therefore, be described by a network metaphor. Different actors, with different, but related, interests, perspectives, norms and values interact to produce the phenomena we study (Klijn 1997).

These three notions imply that the generation of impacts of evaluation must not be analyzed as the result of characteristics of the evaluation study as such, but as the outcome of processes of collective sensemaking in networks (Van der Meer 1999). With ‘collective sensemaking’ we do not suggest that these processes result (or may result) in consensus. On the contrary, the fact that different actors have different histories and, at least in part, different contexts implies that consensus is virtually impossible. However, we derive from the presuppositions above the idea that sensemaking and (changes in) behavior are not produced by actors in isolation either. Hence, we argue

that the impact of evaluation is not so much dependent on the question of how many actors read the report and what they individually make of it, but on when and how it contributes to changes in social patterns. The way in which evaluations are linked to or embedded in existing social and cognitive patterns will be an important determinant in this connection. We should emphasize, however, that 'linked to' and 'embedded' in is not the same as perfectly fitting in. Any evaluation that is to have some impact should (in the perception of at least some actors) have something new or unexpected to say.

Towards constructive evaluation arrangements

In this section we try to come to guidelines for realizing constructive evaluation arrangements in which the contradicting demands of accountability, learning and cooperation are met. We depart from a constructivist evaluation perspective (Majone 1989; Guba and Lincoln 1989; Fischer and Forester 1993) since it regards policy making as a process that can be shaped and given meaning in an interactive way by various interested parties. We regard this type of evaluation as (viz. Guba and Lincoln 1989):

- a sociopolitical process where interests can play a role;
- a cooperative process between evaluators and evaluated;
- a learning and educational process;
- a continuous, recurrent process;
- an emergent, slowly forming process;
- a process that creates common images and collective actions.

Linking typical problems with evaluation in multi-actor policy processes with insights in how evaluation impact is generated, we conclude that a key question is: what do evaluation activities and results mean to different actors involved? Although partial or tentative answers to this question do not dictate how evaluation should be done, they help in thinking about what evaluation activities may be sensible for what, for whom and under what conditions. It may also help to think about how different evaluation activities are or may be related to each other.

Our idea is not that one coherent structure of evaluation activities can be build in a complex policy field such as spatial planning. Different actors will always have (partly) different evaluation interests, involving different criteria, different boundary conditions, different conceptual and theoretical underpinnings, etc. What might be possible, however, is to organize more interrelatedness (connectivity) between evaluation activities and thus to find ways to enhance the utilization and impact of evaluation and to manage the types of problems we identified. Eventually, of course, such interrelatedness can only be produced by the actors themselves. They will have to construct an evaluation arrangement (or set of arrangements).

We now suggest a number of principles for the construction of an evaluation arrangement in a complex multi-actor policy field. We distinguish three different kinds of principles: (A) with respect to single evaluations, (B) with respect to linking of evaluations, and (C) with respect to an arrangement of interconnected evaluations.

A. Principles with respect to single evaluations

1. *Design evaluation projects carefully, i.e. pay tribute to generally accepted technical and methodical requirements, but do also pay attention to the embeddedness of the evaluation in the policy field.*

Care for embeddedness asks that due attention be given to problem definitions and knowledge interests of other actors. This is not to say that (all) other actors should agree with all elements of the evaluation design, but that extent to which a particular evaluation is a basis for communication between interdependent actors, is an important quality parameter. So, perspectives of other actors should be pictured and the question whether and how they be taken account of are essential questions in evaluation design. Another point in relation to embeddedness is the role and position of the evaluator. What does his/her (in)dependence or approach mean to different actors?

2. *Timing of evaluations should be linked to the time schedule of policy processes at different levels*

The initiator of an evaluation should ask himself and other relevant actors when evaluation results will have a relatively high probability to be seriously considered by policy makers or implementers. Generally this will be in phases in which (new) policy measures are (still) considered. It must be realized, however that policy processes may not be in the same phase for different policy actors although their policy actions are interdependent. Apart from dealing with different preferences with respect to timing

there is the question what timing of certain evaluations may support the collaboration and coordination of actions of different actors.

3. Distinguish and link static and dynamic elements in evaluations

In dynamic policy processes it does not seem very sensible to stick only to predefined goals and criteria, even if the primary goal of an evaluation is to control or judge a policy and its implementation. This is so because such a focus would not take external change or internal learning processes into account. Still, there is something to be learned from evidence about the extent to which initial goals are realized, both about goal realization and about goal setting. But accountability may also refer to the measure to which a policy is adapted to changed circumstances, new goals generated in the process or complications observed in collaborative production of policy results.

B. Principles with respect to linking of evaluations

4. Link internal (self-)evaluations to external evaluations

Above we pointed out that there may be a serious tension between external evaluations and internal evaluations if we focus on learning. By making combinations, e.g. by linking a self-evaluation to an external visitation, this tension may be made productive. If the focus is accountability, this procedure may help to prevent that external evaluators draw conclusions that are not considered justified by the agency evaluated. If the focus is learning or improving collaboration, the procedure may enrich internal sensemaking by ‘independent’ observations. Thus both functions may be effectively combined

5. Link quantitative evaluation, qualitative evaluation and trend analysis

Hard quantitative data on policy realization and effects may be very illuminating. To fulfill a function in accountability or learning processes, however, they will have to be made sense of, typically by different actors at the same time and often in mutual consultation. When there are related reports on general (‘autonomous’) trends and qualitative assessments (e.g. on how citizens do experience the policy concerned and its effects), such sensemaking (discussions) will be facilitated (see also Abma, 2001).

6. *Distinguish and link product evaluations and process evaluations*

Some policy actors may be (or feel themselves) primarily responsible for management of policy processes (such as central government to a large extent in Dutch spatial planning), while others are also (or primarily) responsible for the content (policy products). But since processes and products as well as the actors producing them are interdependent in multi-actor processes, organizing connections between the two types of evaluation will put both in perspective and facilitate consultation and coordination between policy actors having different responsibilities.

C. Principles with respect to an evaluation arrangement

7. *Link evaluations performed at different levels*

There will be evaluations at central level next to those on regional and local level. There will also be relevant evaluations at the European level. But they can be linked in a couple of ways. It is possible to have mutual consultations about attention point of one level that should be included in evaluation at an other level in order to enhance the usability of the evaluation, to facilitate the communication and cooperation between levels and, more specifically to generate information that is relevant for the interrelatedness of their activities. Another option is participate one way or another in each other's evaluations.

8. *Organize multi-actor evaluations about core policy aspects and about joint responsibilities*

Where eventual policy results cannot be attributed to the policy actions of one single actor, mutual evaluations with ‘co-producing’ actors seem very sensible, especially where core aspects of spatial planning policies are concerned. An interesting aspect of forms of multi-actor evaluation is that it implies consultation between the actors involved. So, even where actual policy coordination is a stubborn matter, joint evaluation may facilitate and solidify cooperation, apart from its substantive findings. Joint evaluations can create stability and durability in cooperation, which in turn can result in the leap of faith of taken up joint accountability for policy-making processes.

Concluding remarks

The principles we proposed in the preceding section are intended to stimulate thinking about how connections and connectivity between different forms of evaluations in a complex multi-actor policy field can be improved. This thinking can be done by different actors at the same time and requires (at least where rules 4 to 8 inclusive are concerned) consultation and participation. In other words: cooperation seems vital to preserve and combine the other two functions of evaluation. A limited number of genuine multi-actor evaluations may constitute central points in the total evaluation arrangement, surrounded by many more or less – and in different ways – connected

evaluations, most of which are designed in a way that they may be at least of some interest to at least some other actors.

If principles such as these are applied, resulting in some arrangement, then the key question for its viability is whether it offers enough for different actors to invest in it. For coordination and connectivity can be arranged, but it only materializes if actors live up to it in their evaluation activities and in interpreting evaluation results. Therefore we suggest that evaluation arrangements developed be looked at from the perspective of different actors in turn (perhaps actually with those actors) to explore whether and when they are prepared to invest in it in ways that are attractive to other actors as well. This exercise can be viewed as an *ex ante* assessment of the arrangement. But its conclusions can also help to convince different actors of the attractiveness of the arrangement.

We realize that our argument tends to blur the distinction between policy making and evaluation. We also realize that our suggestions are, at most, a beginning. But we are convinced increasing connections between evaluation activities is essential if evaluation is to contribute to accountability, learning and cooperation in complex multi-actor policy networks.

Footnotes

- (1) The word ‘planning’ may cause some misunderstanding. Actually the Dutch phrase ‘ruimtelijke ordening’ [structuring of space] evades this word.
- (2) In the framework of the Fifth Bill we were asked to reflect on modes of evaluation in which accountability and learning in spatial policies could be supported and combined (Teisman and Van der Meer 2002). The report (in Dutch) can be obtained at request.
- (3) The Bill on Space appears to leave even more discretion to decentralized levels, possibly rendering even evaluation on process criteria difficult.

References

- Abma, T. A. (1996). *Responsief evalueren: discourses, controversen en allianties in het postmoderne*. Rotterdam.
- Child (1972). "Organizational structure, environment and performance: the role of strategic choice." *Sociology* **6**: 1-22.
- Cousins, J. B. and L. M. Earl (1992). "The case for participatory evaluation." *Educational Evaluation and Policy Analysis* **14**(4): 397-418.
- DiMaggio, P. J. and W. W. Powell (1991). *The new institutionalism in organizational analysis*. Chicago, University of Chicago Press.
- Dryzek, John S. 1993. Policy Analysis and Planning: From Science to Argument. Fischer, Frank and John Forester (eds.) *The Argumentative Turn in Policy Analysis and Planning*, Durham/London: Duke University Press.
- Edelenbos, J. (1999). "Design and management of participatory public policy making." *Public Management* **1**(4): 569-578.
- Fischer Frank and John Forester, eds. 1993. *The Argumentative Turn in Policy Analysis and Planning*, Durham/London: Duke University Press.
- Guba, E. G. and Y. S. Lincoln (1989). *Fourth generation evaluation*. Newbury Park, London, New Delhi, Sage.
- House, E. R., & Howe, K. R. (1998). *Deliberative democratic evaluation in practice*. Boulder: University of Colorado.
- House, E. R., & Howe, K. R. (2000). Deliberative democratic evaluation. *New Directions for Evaluation*, nr. 85, 3-12.

- Klijn, E.-H. (1997). Policy networks: an overview. *Managing complex networks: strategies for the public sector*. W. J. M. Kickert, E.-H. Klijn and J. F. M. Koppenjan. London, Thousand Oaks, New Delhi, Sage: 14-34.
- Klijn, E.-H. and G. R. Teisman (2001). Governing public-private partnerships: analysing and managing the processes and institutional characteristics of public-private partnerships. *Public-private partnerships: theory and practice in international perspective*. P. Osborn. London, Routledge: 165-186.
- Leeuw, F. L. and P. J. Rozendal (1994). Policy evaluation and the Netherland's government: scope, utilization and organizational learning. *Can governments learn? Comparative perspectives on evaluation and organizational learning*. F. L. Leeuw, R. C. Rist and R. C. Sonnichsen. New Brunswick/London, Transaction Publishers: . 67-86.
- Majone, Giandomenico. 1989. *Evidence, Argument, and Persuasion in the Policy Process*. New Haven: Yale University Press.
- March, J. and J. Olsen (1989). *Rediscovering institutions*. New York, Free Press.
- Mayne, J. (2003). Results-Based Governance: Collaborating for Outcomes. *Collaboration in public services: the challenge for evaluation*. A. Gray, B. Jenkins, F. Leeuw and J. Mayne. New Brunswick (USA) and London (UK), Transaction Publishers: 105-130.
- Mead, G. H. (1962 (1934)). *Mind, self and society from the standpoint of a social behaviorist*. Chicago, London, University of Chicago Press.

- Mulder, H.-P., G. Walraven, A. d. Groot, F. S. Terpstra, P. Rozendal, O. Delsman, A. v. Loon, R. Venderbosch and F. L. Leeuw (1991). "Gebruik van beleidsevaluatieonderzoek bij de rijksoverheid." *Beleidswetenschap* **5**: 203-228.
- Patton, M. Q. (1997). *Utilization-focused evaluation*. Thousand Oaks, London, New Delhi, Sage.
- Provus, Malcolm N. 1971. *Discrepancy evaluation*. Berkeley, CA: McCutcheon.
- Rist, R. C. (1994). The preconditions for learning: lessons from the public sector. *Can governments learn? Comparative perspectives on evaluation and organizational learning*. F. L. Leeuw, R. C. Rist and R. C. Sonnichsen. New Brunswick/London, Transaction Publishers: 189-205.
- Scharpf, F.W., (1997), *Games real actors play; actor-centred institutionalism in policy research*, Boulder Colorado: Westview Press
- Schwandt, Thomas A. 1984. *An examination of alternative models for socio-behavioral inquiry*. Unpublished dissertation, Indiana University.
- Selznick, P. (1984 (1957)). *Leadership in administration*, University of Berkeley Press.
- Shulha, L. M. and J. B. Cousins (1997). "Evaluation use: theory, research and practice since 1986." *Evaluation Practice* **18**(3): 195-208.
- Stake, R. E. (1983). Program evaluation, particularly responsive evaluation. In G. F. Madaus, M. Scriven, & D. L. Stufflebeam (red.), *Evaluation models*, Boston: Kluwer-Nijhoff, pp. 287-310.
- Stufflebeam, Daniel L. 2001. Evaluation Models. *New Directions for Evaluation* **89**: 8-98.

- Teisman, G. R. and F.-B. Van der Meer (2002). *Evalueren om te leren: naar een evaluatiearrangement voor de Vijfde Nota RO*. Rotterdam, Erasmus Universiteit.
- Van der Meer, F. B. (1999). "Evaluation and the social construction of impacts." *Evaluation* 5(4): 387-406.
- VROM (1988). *Vierde Nota over de Ruimtelijke Ordening*. Den Haag, Ministerie van VROM.
- VROM (1992). *Vierde Nota over de Ruimtelijke Ordening Extra*. Den Haag, Ministerie van VROM.
- VROM (2004). *Nota Ruimte: Ruimte om te ontwikkelen*, Den Haag,. Den Haag, Ministerie van VROM.
- VROM, M. v. (2001). *Ruimte maken, ruimte delen: Vijfde Nota Ruimtelijke Ordening 2000/2020*,. Den Haag, Ministerie van VROM.
- Weick, K. E. (1979). *Social psychology of organizing*. Reading (Mass.), Addison-Wesley.
- Weick, K. E. (1995). *Sensemaking in organizations*. London, Sage.
- Weiss, C. H. (1980). "Knowledge creep and decision accretion." *Knowledge: creation, diffusion, utilization* 1(3): 381-404.
- Werkgroep Vijfde Nota Ruimtelijke Ordening (1999-2000). *Notie van ruimte: op weg naar de Vijfde Nota Ruimtelijke Ordening*. Den Haag, Tweede Kamer 27210.
- Woltjer, J. W. (2000). *Consensus planning: the relevance of communicative planning theory in Dutch infrastructure development*. Aldershot, Ashgate.