THE LEARNING EVALUATION
A Theoretical and Empirical Exploration

JURIAN EDELENBOS
ARWIN VAN BUUREN
Erasmus University Rotterdam, the Netherlands

In this article, the authors theoretically and empirically explore the concept of learning evaluation. They shed light on the positioning of the learning evaluation amid scholarly work on evaluations. Moreover, they describe the learning evaluation in practice in the Netherlands by going into a specific project called the Stimulation Program on Citizen and Environment. The theoretical and empirical quest gives insights into the problems with and possibilities of the learning evaluation. They think that their experiences can help the further development of theory about learning evaluation as well as aid in the practice of such evaluations.

Evaluations are part of everyday life in government. There are various types. In the Dutch policy context, scholars have recently been considering the "learning evaluations" that focus on trying to continuously improve projects as they unfold during the implementation process instead of focusing on assessing success or failure of projects after they are realized. Assessment is a component of this type of evaluation, but there is also an element of learning. The learning evaluation method is not without its problems. It imposes various demands on evaluators and generates substantial role conflict for them. Hence, there is every reason to consider this evaluation type.

In this article, we first shed light on the position of the learning evaluation amid the avalanche of scholarly work on evaluations. We will be able to position this type of evaluation better by using a learning evaluation project that we analyzed ourselves. It will become clear that learning evaluations cannot be easily placed in a number of classic dichotomies. It will also become clear that in practice, learning evaluations generate tensions. We think that our experiences can help the further development of theory about learning evaluation as well as aid in the practice of such evaluations. We realize, however, that we report from only a single case study, and therefore we need to be

careful in generalizing insights gained from this case study to general conclusions. The learning evaluation we executed is a unique experience, although it has elements of different existing evaluation methods. We come back to this point in our concluding section. We try to make our experiences beneficial for other evaluators by formulating program characteristics that are suited for a learning evaluation and skills needed to carry out this type of evaluation.

Together with a number of researchers, we evaluated the Stimulation Program on Citizen and Environment (SPCE) of the Department of Environment of the Dutch Ministry of Housing, Spatial Planning and the Environment (abbreviated in Dutch to VROM) using a learning evaluation approach (Edelenbos et al. 2003). This evaluation research was an exploratory expedition for us in the field of evaluation research. We did not pursue a classic evaluation (at a distance from the object, not mixed with practice) but a learning evaluation (the content of which was unclear at the start of the project).

The fact that this research was an exploration resulted in a specific setup of the article. Because we had not considered in detail what the learning evaluation for this project would exactly look like, we do not start with an extended conceptual treatment of it. The objective of the article is, rather, to arrive at a more explicit conceptual image of this type of evaluation through an empirical quest. For this reason, we begin with a brief description of the trends and developments in the field of evaluation. Next, we describe the quest we undertook during the shaping and execution of the evaluation of the SPCE. We address our learning experiences, and we summarize the characteristics of the learning evaluation and link it to existing types of evaluation. We end with some conclusions about the applicability of this type of evaluation and the specific skills evaluators must have to do such an evaluation.

TYPES OF AND DEVELOPMENTS
IN EVALUATION RESEARCH

TYPES OF EVALUATION

Time and again, politicians and administrators ask for an evaluation of policy they or others initiated. These questions usually find their way into evaluation studies and reports. The academic world also has interest in this phenomenon. Many authors have studied evaluation (Stufflebeam and Shinkfield 1990; Madaus, Scriven, and Stufflebeam 1983; Guba and Lincoln 1989; Rossi and Freeman 1999; Patton 1997).

The scholarly reflection on evaluation research has resulted in an enormous growth of categorizations of types of evaluation (see for an extensive categorization Stufflebeam 2001). A distinction is often made between formative (*ex ante*) and summative (*ex post*) evaluations (Scriven 1991). Another distinction is made according to the person who or the body that evaluates (Shulha and Cousins 1997; Wadsworth 2001). Internal evaluations are conducted by those who are to be evaluated; they determine questions and methods as well as judge the results. An external evaluation is conducted by an independent or supervisory body. A third distinction is inspired more by philosophy of science and considers the manner of evaluation. Rational approaches are then distinguished from more constructivist approaches. Rational evaluations emphasize the determination of the degree to which policy programs meet the targets set earlier by a hierarchically superior actor (Tyler 1942, 1966; Provus 1971; Steinmetz 1983). In contrast to this monocentric model, there is the pluricentric model. Constructivist evaluation considers more 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; van der Knaap 1995).

DEVELOPMENTS IN EVALUATING

A number of trends can be perceived in the development of evaluation research. It is tempting to place all the evaluation research of the past decades into one category as "classic," but that fails to do justice to the rich variety. Still, there are some common elements in the mainstream of traditional evaluation research (van der Knaap 1995).

Thinking about evaluation changed in the 1980s in the context of a shift in thought about government and how policy ought to be made and implemented. The classic view of government as an actor that steers society was increasingly abandoned. Other conceptualizations of the policy process became fashionable. As a consequence, perspectives on policy evaluation changed as well.

We first saw a call for more democratic forms of evaluation in literature on participative, constructivist, and responsive types of evaluation (e.g., Scriven 1991; Stake 1983; Schwandt 1984; Guba and Lincoln 1989; Abma 1996; House and Howe 1998, 2000; Abma and Stake 2001). To various degrees, evaluation studies were conducted together with those involved. A first step was to involve the actors in the execution of the policy program (Patton 1997), such as employees, but also the customers or consumers. Next,

stakeholders, who are not customers but individuals and organizations that have an interest in the evaluation, were included (Guba and Lincoln 1989). Finally, participation included all interested citizens (what is often called "community care"). In this instance, one can speak of deliberative evaluation (House and Howe 1998, 2000).

Central to this participative ideal was the notion that evaluation research was shaped and executed in communication with the participants. The choice of participants depended on various factors. These included practical conditions, such as the desire of the principal, the objective of the evaluation, the characteristics of the policy project to be evaluated, and so forth.

Participative types of evaluation fit in the demystification of both policy-making and policy evaluations. Constructivist notions have clearly overshadowed the rational perspectives (Majone 1989; Fischer and Forester 1993; van der Knaap 1995). A central notion is that people themselves give meaning to the world around them. Intersubjective agreement is the best one could hope for with regard to knowing the world. Each representation of reality is normative, and neither policy maker nor evaluator has a prerogative on the truth. To execute a meaningful evaluation, it is crucial to have agreement between various parties.

In this demystification, a change from a judgmental to a more open and investigative evaluation style was appropriate. These two approaches differ in many ways (see Table 1) (Wadsworth 2001, 49).

The conclusion that policy is not made in closed administrative circles but in networks consisting of many actors who are active and will continue to meddle with the policy after political decision making stimulated thinking about models other than the classic top-down types of evaluation. These are highly ex post by nature: A program is assessed in retrospect for its outcomes. Thus, politicians and administrators can be held accountable for their deeds. Acknowledging the complexity of policy processes generates thought about milder types of evaluation focused on improving the functioning of programs. Fitting for this are more *ex durante* evaluations: closely tracking a program during execution and providing reflections that might lead to adaptations.

Finally, we can view the evaluation from the user perspective. Much evaluation research is utility oriented (Patton 1997). The desire of the user and his or her worldview are an important guiding factor in the structure of evaluation studies. Evaluations must be responsive (Rossi and Freeman 1999). This sets specific demands on evaluation research. "Implementation of a utility-focused, feasibility-conscious, property-oriented, and accuracy-based evaluation requires situational responsiveness, methodological flexibility,

TABLE 1: Inquiry and Audit Review Evaluation Approaches

Inquiry Evaluation	*Audit Review Evaluation*
Inquiry: to seek	Audit: to check
Starts with the questions: How are we going? How is this service or activity going? Is it working? In what ways? What do we think of this service? What is its value?	Starts with the questions: Have we done what we set out to do? Is this service, activity, meeting its objectives?
Asks the comparative questions: What are we doing and is that good or bad?	Asks the comparative questions: What did we set out to achieve, and what are the signs we have done this?
Then asks problem-posing and problem-solving questions.	Then asks the gap-filling and "irrelevance"-eliminating questions.
Implies asking: What are the needs?	Implies already assuming what the needs are.
The questions are "opening up" questions, implying the need to build theory from diverse sources.	The questions are "narrowing down" questions, implying the need to test theory from preexisting sources.
Starts with immediate or obvious "problematization"; leaves nonproblematic as taken for granted.	Sets out systematically to problematize all possible activities.
Examines practice to be able to extract assumptions and intentions; can then develop new and improved evaluative criteria.	Examines practice in the light of objectives (applies known evaluative criteria).
Requires a questioning, intuitive, observant (interpretative), inquisitive, imaginative, speculative, and creative mind.	Requires a systematic, orderly, observant (monitoring), fastidious, highly organized, analytical mind.
Relies on who the inquirer/s is/are.	Relies on the quality of previously agreed-on goals, objectives, and aims (and level of consensus previously reached).

multiple evaluator roles, political sophistication, and substantial doses of creativity, all elements of utilization-focused evaluation" (Patton 1997, 17).

These trends are closely related. A more participative angle has everything to do with enlarging the user value of studies. Rossi and Freeman (1999) argued that a participative evaluation process can lead to better understanding of the results and a larger awareness that they are valid, reliable, and convincing, thus leading to more acceptance of them and a larger sense of responsibility to do something with the results.

Many of the trends mentioned above find a place in what Guba and Lincoln (1989) called the "fourth generation" of evaluation research. They regarded this type of evaluation research as

- a sociopolitical process in which 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 with unpredictable outcomes; and
- a process that creates common images.

This type of evaluation also alters the role of the evaluator (Guba and Lincoln 1989; Abma 1996; Edelenbos and van Eeten 2001). In the first generation of evaluation studies, the evaluator was the technician, the analyst. He or she was the one who gave meaning to what he or she observed and tried to acquire an understanding of social reality by making a meaningful reconstruction of it. In the second generation of evaluation studies, the evaluator went one step further. He or she was the one who reconstructed the course of events, made them alive, and shed light on them. The evaluator provided "thick" descriptions (Stake 1983). The third generation of evaluation studies turned the evaluator into the role of judge. The evaluator made an authoritative assessment and told the people in office what they had to do. In the fourth generation, the evaluator was a mediator, a coproducer of social constructions and possible assessments. The evaluator no longer examined events but was a coplayer. Together with the evaluated, he or she tried to arrive at meaningful insights. In doing so, the evaluator taught and attempted to transmit his or her knowledge to arrive at shared knowledge. Finally, the evaluator was a change manager. He or she took responsibility for improving the program and was a key figure in the creation of a "more sophisticated reality" (Guba and Lincoln 1989, 262; Abma 2000).

SPCE

In this section, we describe the execution of the learning evaluation of the SPCE of the ministry of VROM. First, we give some background information on the project. Second, we describe the evaluation process and the results it produced.

BACKGROUND

Creation of the SPCE. The initiation of the SPCE was a direct consequence of a political motion in the 2002 budget proposed in the Dutch House of Commons (Second Chamber) by the MPs Feenstra, Klein Molenkamp, and Augusteijn-Esser (House of Commons 2001-2002), arguing that citizens had to play a greater role in the development of the Dutch environmental policy.

The aim of this motion was to involve citizens more in a coproducing role in making environmental policy. This objective was strongly emphasized and fit a broad political desire to bring government functioning closer to the citizen. The assumption was that trust in national environmental policy would increase when government showed that it believed the "citizen agenda" to be important. In addition, environmental policy would gain in legitimacy and effectiveness when it took the priorities and behaviors of citizens into account.

The Feenstra amendment was adopted by the then minister of VROM, and a line was added in the budget for a multiannual program aimed at shaping the involvement of citizens in environmental policy making: the SPCE.

The four "citizen and environment" projects. The SPCE was elaborated at administrative levels of the ministry of VROM where two main objectives were set. The first was to strengthen "citizen orientation" in environmental and sustainable policy and focused on policy makers. They had to learn to take into account the effects of policy measures on the daily lives of people.

The second was to involve citizens in policy making: the coproduction of policy. In the context of this program, coproduction included a wide variety of types of citizen participation in policy making and implementation. The SPCE was further elaborated into four projects.

- Projects in the context of policy innovations of the National Environmental Policy Plan 4: In the opening steps of the program, various policy subjects were defined with the various policy directorates that had to be further developed together in the citizen participation project.
- Projects in the contexts of the four (interdepartmental) transitions: sustainable energy, sustainable use of biodiversity and natural resources, sustainable agriculture, and sustainable mobility. Five ministries were involved in this transition policy, and VROM coordinated the policy.
- Projects in the context of the National Strategy for Sustainable Development (NSSD), culminating in the Sustainable Development Action Program. Through the SPCE, the effort was made to strengthen the citizen orientation of NSSD.

- To subsidize projects of societal organizations and institutions in the context of the Citizen and Environment Policy of the Subsidy Regulation for Societal Organizations and Environment (SSOE): Societal organizations could request subsidies for projects focused on helping citizens to help determine the agenda about policy innovations and to bring these to the attention of policy makers, politicians, and the public at large.

To execute the entire SPCE, a program team was formed consisting of four employees of the Department of Environment and three employees from the Institute for Public and Politics, an external body specialized in organizing citizen involvement.

EVALUATION OF THE SPCE

In this section, we describe the process and the outcome of the evaluation. First, we discuss the formulation of the assignment and then the methods used for this evaluation research. Furthermore, we discuss the evaluation process and its results more extensively.

Evaluation assignment. In an early stage, thought was given as to how to shape the evaluation. The program team wanted an evaluation during the program, not when it was finished. They hoped that the evaluation could serve as a state-of-the-art memo for the budget round of 2004. The program team especially wanted an evaluation that enabled them to steer the project during the implementation process. A learning evaluation is aimed at frequently linking back to the program team with information about how the various projects are going. The program team recognized four objectives of the evaluation:

- developing a report for the Second Chamber about the state of the art with regard to the execution of the Feenstra amendment,
- inventorying bottlenecks in the SPCE aimed at improving it,
- analyzing potentials of the approach of the SPCE, and
- acquiring insight in how citizen involvement took shape (orientation of civil servants, participation of citizens, initiation by citizens).

The central evaluation question was this: To what extent has the SPCE been able to get environmental policy more "of, by, and through the citizen," and what recommendations can be given so that the program can better fulfill its main objective?

In this research question, we recognize a classic evaluation research question. The distinctive nature of this research is thus not so much the questions that need to be answered (each evaluation has questions, after all) but more in the manner of how the evaluation is done: frequent linking back to program team and focusing on use of recommendations during the process, as well as the evaluation team considering how the program team deals with these recommendations (their learning capability).

A mix of methods. Evaluation research is characterized by the use of a large number of different methods. In this study, the following methods were used:

- Two random-basis telephone surveys of civil servants, once in January 2003 and once in May 2003: The objective was to test the awareness about the SPCE and to determine the degree to which civil servants were citizen oriented.
- A written survey among projects for the SSOE: In May 2003, project leaders in the SSOE project received a survey to find out about the structure, the execution, and the results of each project.
- Document analysis: This included documents about citizen and environment projects such as research reports, policy memos, letters, project designs, and so forth.
- Interviews: The project leaders were interviewed twice, once in January 2003 and once in May/June 2003. Some members of the program team were also interviewed.
- Observations: Some citizen and environment projects were observed. In addition, the researchers attended the general communication and administrative meetings. Also, some meetings of the program team were attended.
- Learning sessions: An important part of the evaluation involved the so-called learning sessions with the members of the program team. A total of five sessions were organized. The first sessions focused on finding the dominant images in the team about the objectives of the SPCE. Later sessions served to link back information from the evaluation to the program team.

This style of working differs substantially from classic evaluation research. The team of evaluators was not only busy collecting data about how the SPCE and the program team functioned, but it was also clearly involved in advising the program team. Thus, this evaluation has the character of both an assessing and an advising evaluation. This is the first distinguishing feature of a learning evaluation, and it occurred at two levels:

1. Learning at the project level: Specific questions and subjects are evaluated in the context of the project.
2. Learning at the program level: This involves focusing on stimulating learning processes across the boundaries of individual projects (i.e., program team; Edelenbos et al. 2003, 22).

Evaluation process. The evaluation particularly concerned the start of the SPCE: January to May 2003 and the half year prior to that when civil servants developed the program. The final evaluation report was submitted in October 2003. The evaluation process consisted of three main phases:

1. preparation and first exploration (January-February 2003);
2. evaluation, adaptation, and evaluation (March-May 2003); and
3. conclusion, reporting, and interactive writing of recommendations (June-September 2003).

PREPARATION AND FIRST EXPLORATION

There was enthusiasm for this new approach to evaluation research on the part of the principal. An evaluation during the course of the program with explicit attention for intermediate linking back was regarded as very attractive. This provided a fruitful basis for thinking about shaping the evaluation. The evaluators put in a substantial amount of time to clarify the logic of the SPCE and the ambitions held by the team members. A number of learning sessions were needed for this. Yet it appeared that the team members were not in sync: The ambitions varied widely.

Much time went into starting up the evaluation. Although this was important for arriving at a shared "language" and focus, in practice little time was available for this. The projects had already started, and the evaluation of them simply had to start as well. Unfortunately, there was not always enough time (and money) for the learning sessions.

EVALUATION, ADAPTATION, EVALUATION

The research team started its evaluation work more or less simultaneously with organizing these learning sessions. A first round of interviews among project leaders in the various projects that had started and the outcomes of the first "learning sessions" with the program team quickly led to the conclusion that the team had developed many initiatives but that the effects were hardly visible. The projects were very much dominated by the agenda of civil servants; the citizens played a marginal role in this phase. In a learning

session in March 2003, the evaluation team put on the brakes, saying that the SPCE would not achieve its goal of "by, for, and through the citizen" if the current direction was maintained. In short, the program was too fragmented and too few projects were set up on the basis of the citizen agenda. This message led to an intermediate change in direction and an intensification of the SPCE.

The idea of enhancing citizen orientation of policy and policy makers was central when the first projects were started. Intermediate findings through the learning evaluation called attention to the original target of the Feenstra amendment.

In response, the program team wrote a memo for their principal in which it pleaded for a number of intensifications fitting the spirit of the Feenstra amendment. This memo resulted in the commitment to making adaptations in the execution of the program:

- More participation of citizens: Citizens had to have influence in the agenda-setting phase. In the most far-reaching case, the involvement of citizens in agenda formation could lead to a proposal that a subject with high priority and a controversial nature should be settled on the basis of a corrective referendum.
- Emphasis on citizen initiatives: To realize a situation in which more was being done from "the outside in," one would have to create more opportunities for citizen initiatives. Whether citizen initiatives would be picked up by VROM depended on whether they met with barriers at the national level and whether they were meaningful initiatives.

After this written change of direction, the evaluation continued. New learning sessions were organized, the project leaders were interviewed again, a second random survey was sent to Department of Environment employees, and projects were observed.

PHRASING OF AND ACCOUNTING FOR RESEARCH RESULTS

After an intensive period of some 3 months, with intensive learning sessions, research, and interaction, the evaluation team withdrew to write the first concept of its report. The first concept was discussed with the program team in July 2003.

It met with much resistance, and there was clearly a "frightened" response. The evaluators had continuously thought along with the program team, had then been "invisible" for a while, had taken a more distant stance, and "suddenly" presented confrontational conclusions. This was not expected and certainly not desired. The conclusions were first presented verbally but were

now in print. Obviously, the program team members were set on continuing the SPCE.

Items of conflict were mainly linguistic in nature. The manner of writing was often considered incomplete and blunt. It was decided to give the report meaning in interaction between researchers and program team members. Evaluators delivered the concept versions of chapters, and program members commented on them. This was a difficult process because almost every sentence and word were weighed. The evaluators wanted to persist in their conclusions, but the evaluated wanted to present them in descriptions they could subscribe and share. This was a tensioned episode. In a process of deliberation and collaborative dialogue, the conclusions and recommendations got their final form.

The evaluation report was finished in September 2003, and it was possible it would play a role in the budget negotiations of the ministry of VROM for 2004. The evaluation team explicitly opted for a strategy of mutual phrasing because it considered it necessary that the research results have influence, as well as for the success of the learning evaluation. The report was presented in parliament with little discussion, and it was deemed necessary to continue the SPCE.

THE EVALUATION OF THE "LEARNING EVALUATION"

In this section, we look back at the course of the evaluation process. We apply a "self-evaluation" on an evaluation that we designed and executed ourselves. We also use the comments from the program team that emerged in an evaluation discussion with them in October 2003.

OBSERVATION, REFLECTION, AND (RE)ACTION

At the start of the learning evaluation, the intention was to frequently link back to the program team and give them the opportunity to adapt the SPCE as a whole or various concrete projects within it on the basis of the evaluation material. The program team indicated that the linking back of the research results could have occurred more frequently. The program team would have liked to see the intermediate results so that it could have continuously seen what happened in the projects.

In the beginning, we invested a substantial amount of time acquiring an understanding of the images of program team members about the SPCE and

their intentions with it. At that time, there was no opportunity for linking back what we had observed through research. When we finally uttered a complaint, it came as relatively unexpected. According to the principal, we should have made more time for linking back our intermediate impressions of the learning sessions. The time taken for collecting observations came at the cost of time for linking back and reflection. Furthermore, the evaluation period was too short (4 months) for effective monitoring of and reporting about adaptations implemented by the program team. When the report had to be concluded, the implementation of the learning items was in full swing.

There has to be some form of coevolution between the steps in the evaluation process and the steps in the program implementation. That means that evaluators as well as civil servants have to be flexible to fit their own work into that of the other.

LEARNING INSTEAD OF DETERMINING THE SCORE

Learning evaluations are aimed more at learning ("what must be done to achieve the objectives") than at determining the score ("you have not achieved this and that objective"). An important norm of learning evaluation is the learning capability of actors who play a central role in the implementation of the SPCE: To what extent do they have the will and the opportunity to translate acquired insights in adaptations of the program and the various citizen and environment projects in it?

In the end, the principal picked up many recommendations and lessons from the evaluation study. At first, they were written down in an "intensification" memo. Later, they were also translated by shaping projects from the citizen perspective and agenda and by giving citizens a more active role in the development of policy, not so much a reactive role on policy proposals from civil servants. Furthermore, a summary of the evaluation report was sent to the Second Chamber. The learning evaluation is primarily aimed at making sure that recommendations, conclusions, and advice have influence: Through a process of giving mutual meaning to observations during the evaluation, the chances are enhanced that they will receive a follow-up. In this case, the principal regards himself or herself as co-owner of the findings.

DYNAMIC TARGETS IN A TURBULENT POLICY PRACTICE

The learning evaluation is highly subject to change. The focus and the locus of the evaluation study change as the program team changes its program as a consequence of new insights from the learning evaluation.

Learning evaluations thus display a highly evolutionary character: The evaluation changes with the developments in the object under study. The evaluation norm of "learning capacity" is generally maintained, but within this, there is continuous recalibration of the effectiveness and efficiency of the program implementation and the projects in it because the objectives and starting points are subject to change during the course of the evaluation. Thus, the attention shifted in the project to a citizen orientation of civil servants instead of an orientation on the citizen agenda. As a consequence, measuring citizen orientation became part of the learning evaluation. During the course of the program, the orientation on citizen agenda emerged again, and thus the focus of the learning evaluation changed with it.

ROLE OF EVALUATORS: TENSION BETWEEN THOSE INVOLVED AND THOSE AT A DISTANCE

During the evaluation, we acted in several evaluator roles. Sometimes, we were judges by passing positive or negative judgment about how the program and its project unfolded, in our opinion. A telling example of this was our attitude in one of the learning sessions in which we indicated that the program had made little progress and was insufficient in involving citizens in the development of environmental policy, so that true results at that moment could not be presented to the Second Chamber.

We also acted as experts and advisors in the area of citizen involvement by pointing out what administrators had to look out for with regard to involving citizens in shaping environmental policy. We especially fulfilled that expert function with regard to various degrees of citizen involvement and the possible methods that could be used.

At other times, we adopted the role of coach/facilitator by actively guiding members of the program team in the various learning sessions in their search for idea formation about and philosophy behind the SPCE. The objective and future of the program were also discussed during these sessions under the guidance of the evaluation team.

At some moments, we were almost regarded as partners and "co–civil servants" who had an equal interest in making the SPCE successful. Sometimes, we had the feeling that we were too servile toward the principal. Initially, we took great pains to think along the lines of the program. This made it more difficult in a later stage to take some distance and formulate conclusions about the functioning of the program in a more neutral stance. At that moment, it was important for us to apply good "boundary work" (Jasanoff 1990), which involves showing involvement and empathy by thinking in

terms of the SPCE and providing practical and useful tips on one hand and maintaining sufficient distance to remain independent and maintain our own professionalism on the other hand.

Finding the right balance was not easy in practice. Discussing the draft report especially created a moment of confusion. Our idea was to have this discussion as a part of a learning evaluation. This was a reflective moment to us, a moment to link back information to the program team. But the program team experienced this as a "scary" moment: All of a sudden, the report was there. The program team indicated that at some point they had lost sight of us. As evaluators, we had dropped back to a classic, distant role: taking a break to write down our observations. They heard little from us during that period, whereas we had almost daily contact about the project before this time. Once the draft report was done, the program team perceived it as a formal moment to be critical as principals about the analysis and conclusions of the researchers. We regarded the discussion of the draft more as a step in a learning evaluation, as a moment of searching together for the right content and meaning in the report. The program team was very critical and defensive and adopted a formal attitude in place of the open and receptive attitude that they had during the learning evaluation period. We should have communicated better about this function of the meeting.

CONCLUSION AND DISCUSSION

In this section, we attempt to draw some conclusions about the learning evaluation and attempt to position it in the vast landscape of evaluation studies.

LEARNING EVALUATION AS ACTION RESEARCH

Evaluation is an activity related to practice in which it accumulates experiences itself. In the process, reflection and action in the context of a practical challenge are continuously involved with each other. Evaluating is not so much an effort to achieve an objective that was set previously, according to a set track, but it is pursued from an open attitude, departing from general research goals and being creative with insecurity and unpredictability in a continuously changing environment (concerning both the existing and desired situation). Evaluation is a process that is characterized by defining, time and again, the practical circumstances and by developing, time and again, ideas and insight in interaction with these practical circumstances. Evaluating is

then a process of trying and adapting, of acting and interpreting, of learning on the job, and of learning by doing.

A learning evaluation is a form in which users (the evaluated) and executors of evaluation (evaluators) shape the evaluations in close interaction and consultation. An important element is the existence of frequent cycles of observation, conclusion, and (re)action. Observation and conclusion are not the end of an evaluation. A dominant element in the role of an evaluator is to be a "reflective practitioner" (Schön 1983).

The evaluator is closely involved in the process of policy making and, in a way, even a part of it. The evaluator does not relate to his or her environment in an impersonal manner. In uncertain and unique situations, for which standard solutions are not available, the evaluator needs to contribute to this policy context in which he or she is part of the policy practice in a reflexive way. The evaluator is in constant interaction with the actors he or she is evaluating. The actors must respond to the intermediate conclusions after which the evaluator will determine their effects. Alkin (1990, 74) called this "situated responsiveness." This makes the learning evaluation a type of action research (Buijs 2004). According to Greenwood and Levin (1998, 75-76), action research has several features:

- It is context bound and focused on practical problems;
- participants and researchers generate knowledge in mutual dialogue;
- they use the diversity of experiences and capacities in the group of investigated actors to enrich the research process;
- the images constructed in the research process lead to action, and the reflections may lead to the construction of new images; and
- the reliability and validity of action research can be measured by seeing to what extent action, based on research, really resolves problems.

Action researchers are clearly oriented on helping the policy practice they investigated and making a contribution to its improvement together with the actors involved (Stringer 1996; Greenwood and Levin 1998; Wadsworth 2001, 52).

A MULTIVOICED CONTEXT

In the learning evaluation, the evaluator serves several target groups simultaneously: first, the principal (in this case, the program team) but also the other parties involved (in this case, the project leaders in the civil service, involved societal groups, citizens who are to be more closely involved in environmental policy, and the Second Chamber).

This multivoice situation may easily give rise to role conflicts. The evaluator must indicate that he or she does not wish to be claimed exclusively by one specific target group. His or her role also involves bridging opposition and playing the role of negotiator (Guba and Lincoln 1989), and the evaluator must not forget that his or her own problem definition is an object of negotiation. This creates a field of tension (Weiss 1998). On one hand, the evaluator must strive as much as possible to do justice to the plurality of reality definitions of those involved. He or she must invest energy in achieving consensus over his or her evaluation approach, methods, and outcomes. On the other hand, the evaluator must maintain distance, guaranteeing independence to present his or her conclusions with some authority. Finding the right balance between the different roles (assessor, facilitator, coach, negotiator, etc.) is thus a key skill for an evaluator.

THE MULTIFUNCTIONALITY OF A LEARNING EVALUATION

A learning evaluation places less stress on achieving the objectives determined by the principal. After all, a variety of objective definitions play a role, and they also develop over time. Part of the learning evaluation is to help bring this variety of goals and ambitions to the surface and interconnect them. Distance is kept from the objectives set because a central element of the learning evaluation is that insights can change that might lead to an adaptation of ambitions during the policy process.

Hence, central to the learning evaluation is not reckoning of earlier objectives but the learning and responsive capacity, the ability and desire to act according to acquired new insights. In terms of this case, this means the ability of the program team to translate evolving insights (acquired through the learning evaluation) in an adaptation of the program (objectives, approach, implementation, etc.).

However, the rational evaluation does not lose any force. In this case, there is a vertical accountability to the ministry of VROM and the Second Chamber. The Feenstra amendment provided an important frame of reference for shaping and implementing the SPCE in which starting points and objectives were formulated. Appreciation of the program depended on the degree to which VROM took these starting points and objectives into account and the degree to which they were met. Here, we clearly see a tension between classic evaluation research and the learning evaluation. On one hand, the program team must be held accountable for its results. On the other hand, the evaluators try to improve the program while it unfolds. Obviously, various translating moments took place in VROM during the implementation of the

program: New elements (objectives and starting points) were added and emphasized, a "citizen orientation" of civil servants especially became more central and must thus be included in the learning evaluation. Other than an ex post evaluation, the learning evaluation concerns the initial objective as a given for the implementation of the evaluation. Objectives are reinterpreted or developed in the course of a program; they are worth being evaluated. The task of the learning evaluation is making other objectives explicit and creating openings for adding these new objectives to the initial political-administrative objective. Thus, the appreciation of the program can be interpreted in a wider scope.

The learning evaluation is thus a hybrid evaluation type in which elements of other evaluation types mentioned in the literature can be found. A learning evaluation is

- a rational evaluation (Tyler 1966; Provus 1971) because it is focused on assessing whether objectives have been achieved and/or whether objectives reformulated and adapted during the process have been achieved;
- a constructivist evaluation (Majone 1989; Guba and Lincoln 1989; Fischer and Forester 1993) because it regards policy making as a process that can be shaped and given meaning in an interactive way by various interested parties;
- a responsive evaluation (Stake 1983; Abma 1996, 2000; Abma and Stake 2001) because it is focused on continuous information exchange between users of the evaluation and departs from an open and flexible structure that is refined during the process;
- a participative evaluation (House and Howe 1998, 2000) because it strives to give participants a voice in the structure and execution of the evaluation, especially with regard to information provision to arrive at a more balanced analysis of the object of evaluation; and
- a utilization-oriented evaluation (Shadish 1995; Shadish, Cook, and Leviton 1991; Patton 1997) because it is focused on the highest possible effect of recommendations, conclusions, and advice for policy practice.

Above all, the learning evaluation has its own added value because it is focused on the degree to which objects of an evaluation display adaptive and learning capability. In addition, a unique feature of the learning evaluation is that it is not conducted before the start of a program (ex ante) or after its conclusion (ex post) but during the execution of a (policy) program (hence, ex durante). When the learning evaluation succeeds in combining the various types of evaluation, it provides an arrangement that is full of tension and yet valuable for practice.

Programs to be evaluated need to incorporate certain features to be considered for a learning evaluation method. We think the following characteristics are relevant:

- Willingness to learn and to adapt the program: The principals of our evaluation research explicitly announced that the evaluation research had to give input for readjusting the program. This open and learning attitude provided a precondition for carrying out the learning evaluation; otherwise, we did not dare doing it.
- Program in developing phase: An important precondition for executing a learning evaluation is that the program under evaluation is in its developing phase. The program we studied was under construction and new for the civil servants of the government agency. They needed the assistance to make this program successful.
- Continuous feedback to the principal and participants: The evaluation process was set up in a way that evaluation information was not only provided at the end of the evaluation but also, and especially, throughout the execution of the program. This evaluation feedback needed to be organized thoroughly by recognizing opportunities to make the feedback effective. In our case, we knew that the program team was signing up for an update letter for the management team. For us, this was the right time to give feedback with a reasonable chance of impact.
- Sense of urgency: There was a sense of urgency for the learning evaluation. Continuation of the program depended on positive signals and an official "go" from the Second Chamber. The learning evaluation was input for the official assessment in the Second Chamber. So, the evaluation has to have authority to become effective.

Effective execution of the learning evaluation also requires certain skills from the evaluator:

- It is important to be explicit about the cast of the evaluation team. As we mentioned earlier, we experienced a role conflict. On one hand, you must show involvement and commitment. On the other hand, you must maintain distance to guarantee (scientific) independence. We ourselves experienced difficulties in performing both roles. We were wearing so many different hats that it sometimes seemed like we needed different heads. We found the solution in distinguishing roles within the evaluation team. Some members got more an evaluation-counseling role with commitment to the program, and other members took more part in the evaluation research at a safe distance from the program. This turned out relatively well in practice. It is important, however, to keep short communication lines between the two groups within the evaluation team.
- It is also important that you keep an open mind and a flexible approach as an evaluator. Sometimes, we thought we had made important observations but

needed to adjust those observations because the program practice changed according to our feedback information. It therefore seemed that we had to readjust and to rewrite our conclusions all the time.

- Moreover, it is important to develop a negotiating style of evaluation (compare also Guba and Lincoln 1989; Stufflebeam 2001; Abma 2000). At times, we had to negotiate between the top-down views of the government agency and the bottom-up views of the citizens. The wishes of the civil servants and the interest of the citizens with respect to environmental policy sometimes seemed contrary. The expert view of the civil servants especially dominated at times; at those times, we stressed that stakeholder involvement meant that those wishes of the citizens needed to be heard and be assessed. In the end, we got ourselves in a mediating process between what civil servants and citizens wanted. Our experience was that the civil servants of the program team were more willing to take the wishes of the citizens seriously than were the civil servants working in the "normal office," who stayed at a certain safe distance from the program.

The learning evaluation is not the simplest way to conduct an evaluation study. However, when the right conditions are present and when it is carefully carried out, it can improve public policy programs. We hope that our report of our experiences with the learning evaluation contributes to further development of evaluations that are carried out on the borders of science and practice.

REFERENCES

Abma, T. 1996. *Responsive evaluating: Discourses, controversies and alliances in the postmodern.* [In Dutch.] Delft, the Netherlands: Eburon.

———. 2000. Negotiated evaluating: "Fourth generation evaluation" of Egon Guba and Yvonna Lincoln. [In Dutch.] *Bestuurskunde* 17:393-403.

Abma, T., and R. E. Stake. 2001. Stake's responsive evaluation: Core ideas and evolution. *New Directions for Evaluation* 92:7-22.

Alkin, M. C. 1990. *Debates on evaluation.* London: Sage.

Buijs, J. M. 2004. Action research: Research strategy for complex policy processes. [In Dutch.] Master's thesis, Tilburg University.

Dryzek, J. S. 1993. Policy analysis and planning: From science to argument. In *The argumentative turn in policy analysis and planning*, ed. F. Fischer and J. Forester, 213-32. Durham, NC: Duke University Press.

Edelenbos, J., M. W. van Buuren, M. van As, M. Duijn, A. Slob, and W. Vermeulen. 2003. *Environmental policy more of, for and through the citizen: Results of the learning evaluation study of the Stimulation Program Citizen and Environment.* [In Dutch.] Rotterdam, the Netherlands: Erasmus University Rotterdam.

Edelenbos, J., and M. van Eeten. 2001. The missing link: Processing variation in dialogical evaluation. *Evaluation* 7 (2): 211-17.

Fischer, F., and J. Forester, eds. 1993. *The argumentative turn in policy analysis and planning.* Durham, NC: Duke University Press.

Greenwood, D. J., and M. Levin. 1998. *Introduction to action research: Social research for social change.* London: Sage.

Guba, E., and Y. Lincoln. 1989. *Fourth generation evaluation.* Newbury Park, CA: Sage.

House, E. R., and K. R. Howe. 1998. *Deliberative democratic evaluation in practice.* Boulder: University of Colorado.

———. 2000. Deliberative democratic evaluation. *New Directions for Evaluation* 85:3-12.

House of Commons. 2001-2002. Motion Feenstra, Klein Molenkamp en Augusteijn-Esser. [In Dutch.] 28 000 XI: 21.

Jasanoff, S. 1990. *The fifth branch: Advisers as policy makers.* Cambridge, MA: Harvard University Press.

Madaus, G. F., M. Scriven, and D. L. Stufflebeam. 1983. *Evaluation models: Viewpoints on educational and human services evaluation.* Boston: Kluwer Nijhoff.

Majone, G. 1989. *Evidence, argument, and persuasion in the policy process.* New Haven, CT: Yale University Press.

Patton, M. Q. 1997. *Utilization-focused evaluation: The next century text.* 3rd ed. Thousand Oaks, CA: Sage.

Provus, M. N. 1971. *Discrepancy evaluation.* Berkeley, CA: McCutcheon.

Rossi, P. H., and H. E. Freeman. 1999. *Evaluation: A systematic approach.* 6th ed. Thousand Oaks, CA: Sage.

Schön, D. A. 1983. *The reflective practitioner: How professionals think in action.* New York: Basic Books.

Schwandt, T. A. 1984. An examination of alternative models for socio-behavioral inquiry. PhD diss., Indiana University.

Scriven, M. 1991. *Evaluation thesaurus.* Newbury Park, CA: Sage.

Shadish, W. R., ed. 1995. *Guiding principles for evaluators.* San Francisco: Jossey-Bass.

Shadish, W. R., T. D. Cook, and L. C. Leviton. 1991. *Foundations of program evaluation.* Newbury Park, CA: Sage.

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 *Evaluation models*, ed. G. F. Madaus, M. Scriven, and D. L. Stufflebeam, 287–310. Boston: Kluwer-Nijhoff.

Steinmetz, A. 1983. The discrepancy evaluation model. In *Evaluation models*, ed. G. F. Madaus, M. Scriven, and D. L. Stufflebeam, 79-100. Boston: Kluwer-Nijhoff.

Stringer, E. T. 1996. *Action research: A handbook for practitioners.* London: Sage.

Stufflebeam, D. L. 2001. Evaluation models. *New Directions for Evaluation* 89:8-98.

Stufflebeam, D. L., and A. J. Shinkfield. 1990. *Systematic evaluation.* Boston: Kluwer Nijhoff.

Tyler, R. W. 1942. General statement on evaluation. *Journal of Educational Research* 35:492-501.

———. 1966. The objectives and plans for a national assessment of educational progress. *Journal of Educational Measurement* 3:1–10.

van der Knaap, P. 1995. Policy evaluation and learning: Feedback, enlightenment or argumentation? *Evaluation* 1 (2): 189-216.

Wadsworth, Y. 2001. Becoming responsive: And some consequences for evaluation as dialogue across distance. *New Directions for Evaluation* 92:45-58.

Weiss, C. H. 1998. *Evaluation.* 2nd ed. Upper Saddle River, NJ: Prentice Hall.

Jurian Edelenbos is an assistant professor in the Department of Public Administration, Erasmus University Rotterdam, the Netherlands. His fields of research and teaching areas are governance, complex decision making, network analysis, public-private partnerships, process management, trust in interorganizational cooperation, transition management, joint knowledge production, institutional innovation, urban planning, water management, and infrastructure. Some of his publications are "Design and Management of Participatory Public Policy Making," Public Management Review *(1999); "The Missing Link: Processing Variation in Dialogical Evaluation,"* Evaluation *(with M. van Eeten, 2001); and "Double-Helix Approach: The Connection Between Analysis and Interaction in Policy Development,"* International Journal for Technology, Policy and Management *(with R. Monnikhof and O. van de Riet, 2003).*

Arwin van Buuren is a Ph.D. student in the Department of Public Administration, Erasmus University Rotterdam, the Netherlands. His research is about knowledge management in controversial policy processes in the field of spatial development and water management. He has published in several Dutch-refereed journals. Some of his international publications are "Conflicting Knowledge: Why Is Joint Knowledge Production Such a Problem?" in Science & Public Policy *(with J. Edelenbos, 2004) and "Monitoring: Functional or Fashionable?"* in Society and Economy *(with D. Kool, 2004).*