

# Effectiveness Evaluation of Environmental Policy

## the Role of Intervention Theories<sup>1</sup>

*Per Mickwitz*

### ABSTRACT

Besides the general increase in evaluations there is in particular a rising demand for evaluations of the effectiveness of policy instruments. This is exemplified in the EU Environmental Council's common position on the proposed 6th environmental action programme, which stresses "ex post evaluation of the effectiveness of existing measures in meeting their environmental objectives". Although effectiveness evaluations are carried out constantly there is huge variations in how the term "effectiveness" is used and how it is implemented in practice. The following three steps will be discussed: Step 1 determining the policy objectives against which to assess its effects; Step 2 finding the effects of a policy; Step 3 combining effects due to the policy with the objectives in order to determine effectiveness. To carry out any of the three steps involves many challenges. These include determining which effects are caused by the policy being evaluated and to what extent, whose goals should be taken into account and how to interpret the combination of objectives and effects. Intervention theories will be discussed as a tool to partly meet these challenges.

**Key words:** Policy evaluation, intervention theory, effects, effectiveness

### 1. INTRODUCTION

Evaluating the effectiveness of environmental policy is stressed so much these days that one sometimes gets the impression that effectiveness evaluation is synonymous with evaluation. Let us start with a few examples.

The 6th Environmental Action Programme for the European Union (1600/2002/EC), which was finally adopted in June 2002, includes one mention of ex post evaluation focusing on effectiveness. Paragraph c in Article 10 states:

"[The objectives shall be pursued by] improvement of the process of policy making through: ... - ex post evaluation of the effectiveness of existing measures in meeting their environmental objectives;"

In the proposal for the third phase of the Finnish cluster research programme named "Ecoefficient Society" one of six themes is directly related to evaluation. Its heading is "Policy instruments and their effectiveness". The Finnish Ministry of Trade and Industry hosted a seminar during the spring of 2002 on the effectiveness and efficiency evaluation of the measures of the national climate change strategy of different sectors.

Sometimes – although not among professional evaluators – the term effectiveness is wrongly used to refer to many positive effects.<sup>2</sup> For evaluators the term effectiveness generally has a precise meaning, it means "to what extent have the intervention's impacts contributed to achieving its specific and general objectives?" (Nagarajan and Vanheukelen 1997, 71)

To carry out an effectiveness evaluation of a public policy thus implies three things: to reveal the objectives of the policy, to identify the effects of a policy and to compare these two. This might sound simple but all three things involve several challenges. This article will elaborate on these challenges and on the role of intervention theories as a specific evaluation tool with which to tackle them.

In order to undertake an effectiveness evaluation as efficiently as possible the three required steps should be taken in a specific order; one should start by uncovering the objectives of the policy and then move on to determining the

effects of the policy. By taking the steps in this order the second step can be limited to effects related to the objectives. If, on the other hand, effectiveness is just one criterion among others the steps should be taken in the reverse order. In this case one would not like to limit one's perception of potential effects to only the effects related to the objectives, but also look for all kinds of unintended affects within, as well as, outside the target area of the policy (e.g. Hildén et al. 2002).

After we have discussed the three steps necessary for effectiveness evaluations we will tackle the question to what degree effectiveness evaluation is just another name for goal-achievement evaluation or if it is something else. This question was already opened up by Michael Scriven (1991, p. 129) in his definition of the term effectiveness, where he states that it is "Often but not necessarily used to refer to the conclusion of a goal-achievement evaluation". Before moving on to the three steps the concept "intervention theory" will be introduced, because it will have an important role in all three steps.

## 2. INTERVENTION THEORIES

Evert Vedung (1997, 301) defines programme theory or intervention theory<sup>3</sup>, as: "All empirical and normative suppositions that public interventions rest upon". An intervention theory is a model "of the micro steps or linkages in the causal path from program [or more generally intervention] to ultimate outcome" (Rogers et al. 2000, 10) on the basis of the detailed assumptions of how the intervention is supposed to work.

Michael Patton (1997) distinguish three different approaches to develop intervention theories: the deductive approach, which is based on the literature; the inductive approach based on field work; and the user-focused approach based on the implicit theory of action of the intended users. Frequently elements from all these approaches can be used simultaneously.

It is common that several intervention theories can be used to describe an intervention, since different groups often have different expectations of a policy, these being the basis of their support of it (Weiss 2000). Often an intervention theory that is based on legislation, on official decisions taken based on it, or on texts used as official

justifications for these can be articulated (Figure 1). Such an "official" intervention theory has a special status among all intervention theories, because it is institutionalized by representatives of the people or their executives in the parliamentary chain of representation and control. Still, the descriptions and justifications of policy instruments are often very general and thus allow for not one but several alternative "official intervention theories", especially on a more detailed level.

Evaluations, and in particular intervention theories, have an important role for an increased mutual understanding among different groups and individuals. By making the assumptions of actors, inputs, outputs and outcomes and especially the causal relationships between these explicit differences in underlying perceptions can be revealed. The potential for learning is especially large when intervention theories are reconstructed through a participatory process. The potential for increased mutual understanding is particularly important in the environmental sector, where belief systems or views of causal relationships are often conflicting.

The role of intervention theories is to describe how the policy is intended to be implemented and function. They are not intended to describe how a policy actually works. The evaluation will then examine to what degree the "real world" resembles the intervention theories, how the intervention has actually been implemented and what effects it has had in practice. There is thus no conflict between linear and simple intervention theories and a "complex world full of paradoxes and contradictions". Since the former is just a tool to examine the later.

An example of an intervention theory is shown in Figure 1, it concerns the Finnish carbon dioxide tax, which came into force from the beginning of 1990. It should be stressed that the figure does not present current energy taxation in Finland, since the taxation has been thoroughly reformed (e.g. Määttä 2000). It is intended to illustrate the actors involved and the assumed mechanisms, when Finland introduced the world's first CO<sub>2</sub> tax in order to reduce climate change and improve the balance of the state budget.

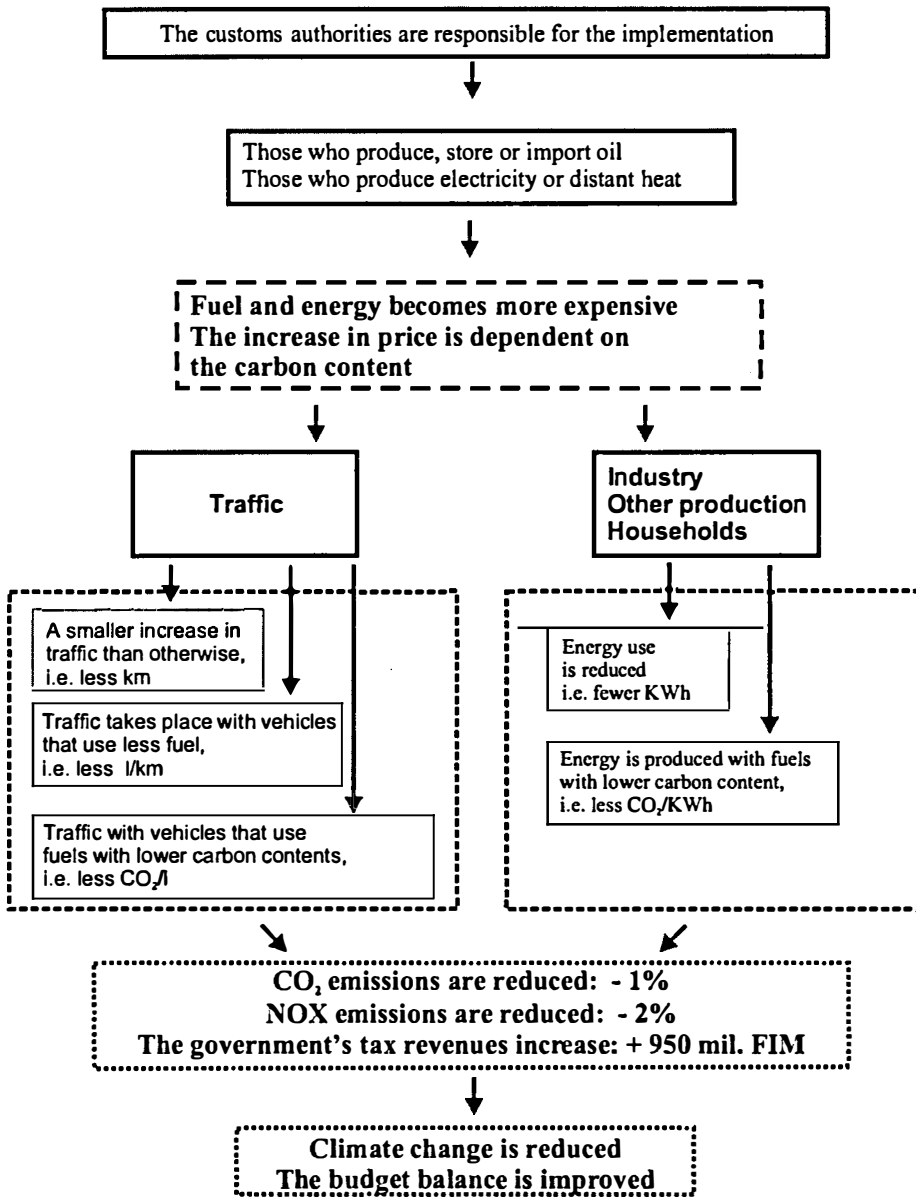


Figure 1. An example of an intervention theory, for the Finnish CO<sub>2</sub> tax introduced in 1990.

**3. STEP 1: DETERMINING THE OBJECTIVES AGAINST WHICH TO ASSESS THE EFFECTS**

**3.1. Challenges involved in determining the objectives**

The first step of an effectiveness evaluation sounds very straightforward; we just have to

find what was supposed to be achieved by the policy. In practice, however, this is often far from simple.

The first question is which goals we should focus on? The obvious first choice is: those goals inherent in the adopted intervention. These are goals that are adopted by democratic institutions, e.g. parliament, government, the Council of Europe or international bodies, such as the parties

to the United Nations Framework Convention on Climate Change. It is therefore very important that these goals be taken seriously and used in evaluations. This first choice is still not always enough or even the only legitimate starting point.

Environmental policies are always based on compromises. These compromises are often the results of long and resource-intensive processes – due to the conflicting objectives involved in many environmental problems – of policy formation.<sup>4</sup> On the one hand this justifies taken into account the goals reached through this process and giving them a special role in the evaluation. On the other hand it is possible that not every interest was represented, or equally represented, thus limiting the scope to the goals considered in the policy compromise could be too restrictive.

The question of whose goals is influenced but not determined by the party commissioning the evaluation. If, for example, an environmental policy has been supported by an environmental nongovernmental organization (NGO) it is of course totally legitimate for them to ask how well their goals have been achieved, i.e. to evaluate the effectiveness against the goals of the NGO. This is the case whether these are a subset of the official goals or even if they are not included in them in any way.

The second question relates to the nature of the goals, often policy goals are very general and vague and thus not very useful from the perspective of an effectiveness evaluation. For example, the objective of the Finnish Waste Act (1072/1993) is "to support sustainable development by promoting the rational use of natural resources, and to prevent and combat the hazards and harm to human health and the environment arising from wastes."

One reaction to the vague goals is that the policy formation process should be developed, so that there would be more precise objectives in the future. For example, in the case of climate change policies the European Commission has formulated guidelines stating "Objectives [should] be described in quantitative terms to the extent possible." (EEA 2001b, 22) There is, however, a clear limit to how specific objectives one can expect. That public policy objectives are multiple, conflicting, and imprecise are not characteristics of bad policies, but inevitable parts of the policy

formation process itself. "The price of agreement among stakeholders is likely to be vagueness, allowing them to fight their battles another day. Policy and program objectives are multiple so as to create a sufficiently broad coalition of support. Contradiction comes in because objectives held by different interest groups ... may well be opposed." (Browne and Wildavsky 1983, 193)

Sometimes it is possible to establish more precise goals by studying the material produced before the intervention was decided on. These include background documents and studies referred to, but also, for example, the records of the discussions in parliament. In other cases the task of focusing the policy objectives may have been delegated to, e.g. a ministry or an agency. In these cases it is often possible to specify goals through reviews of material produced after the decision was taken. However, in many cases even the specified goals may be somewhat vague and qualitative, rather than quantitative. It is often more useful and honest to let this be reflected in the results of the evaluation, i.e. to aim at qualitative statements of the effectiveness rather than artificial quantitative ones.

The third question relates to the time scales of the goals. Often the objectives relate to the final outcomes of the policy, for example the first of the fifteen Swedish environmental objectives is "reduced climate impact" (Swedish EPA 2002). Due to the nature of many environmental processes the time elapsing between action and final outcomes of environmental policy is often very long, frequently decades and sometimes centuries, not all effects can therefore be evaluated at the time when there is a need to know if we are on track. Often it will be necessary to focus the evaluation on outputs and administration and limit the evaluation of outcomes, especially ultimate outcomes. The evaluation thus becomes an evaluation of the prerequisites for effectiveness.

The fourth question relates to the fact that objectives tend to change over time. This is not of course a problem per se; sticking to old objectives in new situations would be a problem, or as Angela Browne and Aaron Wildavsky (1983, 204) put it "It is intelligent to alter objectives to fit resources, to adjust programs to face facts, as well as to fit resources to objectives." For effectiveness evaluation, however, it is a challenge, which objectives should we measure

against? Depending on the purpose of the evaluation – and our resources – one could choose the original objectives, the revised ones or both.

### 3.2. *The role of intervention theories*

When the objectives of a policy are very general they can often be specified through the process of reconstructing the intervention theory of the policy. In some cases when the objectives are very general the process of eliciting intervention theories might reveal implicit specifications of the objectives.

Often intervention theories are useful in order to make explicit which outcome stage the objectives address. While many objectives concern final outcomes, “clean air” or “a magnificent mountain landscape” to take two Swedish goals (Swedish EPA 2002) as examples, other type of goals also exist. Objectives are sometimes stated with respect to outcomes earlier in the chain or even for outputs and there are cases with multiple objectives concerning different stages.

Intervention theories can also be used to tackle the time scale problem, i.e. the situation that arises when an evaluation is demanded before the objectives could have emerged. Intervention theories are tools that can be used to establish the prerequisites for effectiveness. For example, an intervention theory that clarifies the outcomes required for reduced climate change could help in determining the temporal stage on which the evaluation should focus and for which stages the scientific judgment, of e.g. the Intergovernmental Panel on Climate Change, would have to be used.

The intervention theories do not provide much help in resolving the question of whose goals the effectiveness evaluation should be based on. Intervention theories make clear which actors are assumed to carry out the implementation of an intervention and which actors are supposed to change their behaviour because of it, but mostly these actors are known anyway. The challenges of focusing the evaluation on the goals of certain stakeholders and where to find these must be tackled using other tools.

### 3.3. *Problems caused by the goal fixation*

Before moving on to the next step it is also important to be fully aware of the limitations imposed by focusing solely on the goals of a policy. As forcefully argued by Evert Vedung (1997, 49ff.), policies never work exactly as intended, policies tend to have a variety of unanticipated effects, some of which are beneficial while others are detrimental. The main point, however, is that they are not known in advance and thus no specific goals about them can be stated. Therefore if the focus is strictly on effectiveness all these effects will be irrelevant and much that could be learnt from the evaluation will be missed.

Many environmental problems are very complex and characterized by a high degree of uncertainty. It is therefore likely that there will be many unanticipated effects, especially in the area of environmental policy. It is well known that environmental policies that have reduced one problem have often done this by: shifting the problem regionally, e.g. by “higher chimneys”; increasing other problems, e.g. pollution of other substances; or postponing the problem (e.g. Jänicke and Weidner 1995). In an evaluation of environmental policy instruments in Finland increased environmental awareness was seen as partly an unanticipated effect of the developing regulatory regimes and of the expanding environmental specialist education, but also as a cause for policy formulation (Hildén et al. 2002). There were, however, no goals concerning increasing awareness in the 1960s and 1970s.

## 4. STEP 2: IDENTIFYING THE EFFECTS OF A POLICY

### 4.1. *The challenge of determining the effects*

Determining the effects of a policy is often at least as hard as revealing the goals. It involves two activities: first finding out what has happened in the target area and second determining to what degree these changes are due to the policy and not caused by other factors, such as economic growth, general technological development or customer pressures. When the task is only to perform an effectiveness evaluation the effects can be limited to those that are related to the goals of the policy.

For many changes in the environment the key task is to determine the degree to which they are due to the policy instruments. For example, water discharges as well as many forms of air emissions have decreased significantly in many countries in recent decades (e.g. EEA 2002). The questions are, however, to what extent this can be attributed to specific policies and how much other factors such as markets and general technological development are dominant driving forces.

Realising that the impact problem, i.e. to what degree outcomes are caused by the policy, is always very difficult to resolve in social sciences; it can be argued that it is particularly difficult in the environmental policy context. Because environmental problems tend to be very complex, they involve a high degree of uncertainty and in many cases the time elapsing between action and results is very long.

#### 4.2. *Intervention theories*

Intervention theories can be used to determine which outputs, outcomes and causal links to collect data on. The evaluation could focus on the links that are the most crucial for the interventions, the links that are most uncertain or on the links that it is possible to obtain data on. When the evaluation is undertaken the intervention theories will assist in interpreting the results, some assumptions will gain support while others will not, and, more specifically some impacts or lacking impacts will be traced to specific links in the intervention theory. Sometimes the effects of policies can be identified, but the mechanisms are different from those in the intervention theories, then there is a clear opportunity to learn by reformulating the intervention theories of the policy evaluated and maybe also the theories of related policies. Often the intervention theory, or some aspects of it, can be evaluated as such, without data, by comparing its logic to intervention theories of other possible instruments and research findings concerning other policies.

Intervention theories have a special role in evaluations of environmental policy instruments. First they include elements that must be based on science and scientific theories. The fact that there are parts of an intervention theory that

can be based on scientific facts is, of course, a strength compared to most situations in, for example, social policy, where such knowledge is lacking. However, it is also in many ways a challenge: dealing with science requires special skills. Science – especially when it comes to complex ecological issues – often provides no simple and unique theories; and scientific theories can be wrong. Second, due to the long time span of many environmental problems, final outcomes can often not be evaluated in the actual specific situation at hand. As already discussed, in these cases intervention theories can be used to identify the phases to be evaluated instead on the basis of scientific theories or knowledge from other comparable situations.

In a strict effectiveness evaluation the focus is only on the effects and the goals, and thus the implementation process may be deliberately ignored. In such cases the role of the intervention theories is limited to linking the effects to the policy. If, however, the evaluation perspective is slightly broader, the evaluation, even an effectiveness evaluation, is also intended to be used as a learning tool and to improve the implementation process. In such cases intervention theories may be crucial in order to locate the key activities that could be improved.

#### 4.3. *Triangulation*

There is no single clear and universal solution to how one should empirically assess a causal link within an intervention theory. For example, to what degree the immediate outcome depends on the output and how much it is explained by other exogenous variables. Often, however, approaches in which several methods are used instead of only one are appropriate, especially in the environmental policy context. For example, Bartlett (1994, 183) wrote: "Clearly desirable are multiple evaluation, done with a keen appreciation of the strengths and limitations of each approach and a frank recognition of the advantages of others."

Bartlett is by no means unique in advocating multiple methods; often the term "triangulation" is used (e.g. Scriven 1991, 364). Four types of triangulations can be distinguished: multiple methods; multiple sources within one method; multiple analysts; and multiple theories. They

are all important and statistical analyses of data at different levels of aggregation, qualitative analyses of documents, questionnaires and thematic interviews could all be used.

For example, an evaluation of the Finnish environmental policy instruments used for the pulp and paper and chemical industries used statistical analysis in parallel with qualitative analysis of thematic interviews and documents. This data ranged from site level information on waste water discharges, air emission, production and permit conditions to aggregated values for entire sectors. For this data descriptive statistics, such as means, variances etc. were calculated and examined and a variety of graphical analyses were undertaken. In some cases statistical hypotheses testing was performed, for example reductions in phosphorus discharges from mills with a phosphorus limit compared to the reduction of those without limits. A wide range of statistical modelling was undertaken, including time-series analyses of site as well as sector data and probit and logit models of permit contents. The statistical data and analyses were complemented with more than thirty taped and typed thematic interviews, which were coded and analysed. In addition some cases were chosen for which the background documents generated by the legal processes were examined in detail. Since all the analyses were undertaken in parallel the new questions generated by one method were then examined using the other methods, and details required in order to interpret a comment by an interviewee could often be found in the documents or the statistical database. (Hildén et al. 2002)

#### 4.4. *A baseline scenario*

While many evaluations implicitly use scenarios<sup>5</sup>, i.e. at least an assumption of what would happen or have happened without the policy, it is becoming more and more common also to explicitly model scenarios of different environmental policy options. While a variety of scenarios is often used in ex ante evaluations, the most important scenario in ex post effectiveness evaluations is the baseline scenario (some times the terms reference, benchmark or non-intervention scenario are also used). The baseline scenario is a projection, qualitative or quantitative,

of what would have happened without the policy that is evaluated. (EEA 2001a, 11)

One of the most extensive economic ex post evaluations of environmental policy is "The Benefits and Costs of the Clean Air Act, 1970 to 1990" undertaken by the U.S. Environmental Protection Agency (EPA). The assessment is based on a comparison of the actual development of the emissions with a "no-control scenario". In the no-control scenarios the historical emissions are adjusted for the change in the sector's growth predicted without control and specific provisions based on the Clean Air Act. (EPA 1997)

Evert Vedung (1997, 170 ff.) discusses six different designs to illuminate the causal impact: experiments with randomised controls, experiments with match controls, generic controls, statistical controls, reflexive controls and shadow controls. These designs can all be seen as different ways of establishing what Vedung (1997, 167) calls a counterfactual reference outcome. If one interprets the entire difference between the case studied and the control case as the effect of the intervention, then it is not very meaningful to speak about a baseline scenario (although one could say that a scenario is still implicit). Often, however, the differences in the cases are not directly compared, instead the effects of other factors are first subtracted or added. This can be done, for example, through modelling or based on expert judgments. Then the counterfactual development represents what here is called a baseline scenario.

While both scenarios and intervention theories can be useful tools when determining the effects of a policy the questions arise: "what are the differences and what is the relationship between them". There is a clear difference between an intervention theory and a scenario. Although based on perceptions of the real world, often developed through earlier policies, an intervention theory is purely conceptual, an illustration of assumed processes and causal relations. A scenario is an image of how an area might appear or would have appeared without a specific policy, but based on a variety of empirical data and assumptions. The focus of intervention theories and scenarios differs; for intervention theories it is the intended mechanisms while it is the possible "states of the world" for scenarios. Both scenarios and intervention theories are theoretical, since we only have empirical observations on what

happened, not on what might have happened or will happen. Scenarios and intervention theories are related in the sense that scenarios may be built based on the causal relationships and exogenous variables included in the intervention theories and scenarios may be used when intervention theories are formed.

## 5. STEP 3: COMBINING EFFECTS DUE TO THE POLICY WITH THE OBJECTIVES IN ORDER TO DETERMINE EFFECTIVENESS

### 5.1. *The challenges involved in the interpretation*

The final step in an effectiveness evaluation is to combine the results of the first two steps, i.e. the objectives and the effects that are due to the policy evaluated. This sounds fairly easy, but in practice it involves some challenges as well. Assuming that the challenges of the two first steps have been exceptionally well met, that is there are well defined objectives and a good understanding of which effects are due to the policy and to what degree, the following issues have to be addressed while combining these two: "If a goal has been reached is it a sign of a successful policy or a loose goal?" "If the goal has been more than achieved is there a risk of 'overachievement'?" "If a goal has not been achieved is it the policy or the goal that is the problem?"

The first question has to do with strictness of the goals. The result (Hildén et al. 2002, 108f.) alone that most goals of the Finnish environmental programmes have been approached and several have been achieved in due time is not necessary a good indicator of the merit of the policies. The observed effectiveness could be a result of the level of ambition in goal setting. A key civil servant from the Ministry of the Environment has stated that: "In Finland the politically determined goals for environmental administration and environmental protection have generally been realistic or actually cautious. Accordingly, achieving them has seldom imposed big problems." (Ojala 1997, 75) One should therefore not conclude that in cases or countries where goals are not achieved the policy is less successful; it may instead reflect higher ambitions.

Sometimes the policy objectives are stated in such a way that they reflect a minimum or a maximum to be achieved, at other times the aim is stated only as a direction, e.g. the emissions should be reduced. The discovery that the objective has been more than achieved should not necessary be equated with success. There are two basic reasons why overachievement might be a problem. First, if an objective has been negotiated to balance the positive main effect with the negative side effects, then overshooting the target for the main effect could be a sign of poor performance. If, for example, the goal contained in the Kyoto protocol for the European Union to reduce greenhouse gas emissions by eight per cent compared to 1990 reflects what is seen as reasonable - with respect to sacrificed economic growth, higher energy prices, public resources spent on the issue etc. - a reduction of, for example, twelve per cent could be a bad result. Second, if the policy objective is defined as an intermediate outcome the underlying environmental problem may become worse through overachievement. If the ultimate goal is to preserve biodiversity but the intermediate goal is framed as reduced whaling or elephant hunting, it is possible that a total ban on whaling or elephant hunting could result in biodiversity loss in the populations which are next in the food chain or which compete for the same food as these mammals. Similarly, if the final goal is to reduce eutrophication, overachieving a goal for one of the substances behind the problem, e.g. phosphorus, could, according to some theories, make matters worse if nitrogen is not reduced at all.

That goals are not reached and policies are not effective may not always be a problem. As Majone and Wildavsky (1979, 177f.) point out "If both the decision and the execution are... bad, then we can only be grateful that poor decisions are made ineffective by worse actions." Continuing "Outside the static world of programmed decisions, "good" and "bad" take on multiple meanings. In an evolutionary context "good" means "faithful," but interestingly enough, it might also mean "faithless." A faithful translation of an ill-formed policy idea or theory would bring into being all the inconsistencies, inadequacies, and/or unfortunate consequences inherent in the pristine conception." Such a situation would not imply that an evaluation of the effectiveness was redundant.



On the contrary, it would be very important to fully reveal the situation and learn from it.

### 5.2. *The role of intervention theories*

The key to overcome the challenges of combining objectives and effects is to evaluate the goals. That is to assess the relevance of the goals by comparing them with the perception of the underlying environmental problem in hindsight. This implies that the evaluation becomes broader than a mere effectiveness evaluation. Whether the relevance of the policy is explicitly evaluated or not, intervention theories can be helpful in meeting the challenges of combining objectives and effects.

Detailed intervention theories can be useful tools when assessing the strictness of policy goals. This is because by revealing on the one hand the assumptions of what was supposed to take place in order for the goal to be achieved and on the other hand what the actions were that produced the outcomes in line with the objective, something can be said about the ambition. If a policy to reduce air pollution by gasoline taxation was based on an intervention theory that people switch from private cars to public transport but the goal is achieved by the reduced fuel consumption of new cars, the objective is less strict in hindsight than it was when adopted, or the original intervention theory is a tactical one.

Intervention theories can also be applied to judge if overachievement of a goal is a problem or a sign of success. First, if the objective is stated as an intermediate outcome, intervention theories provide the links to the final objectives. Second, if the objectives are reached through weighting of a bundle of positive and negative effects, all the anticipated ones could be revealed through intervention theories.

There are cases where the relevance of the goals has been reassessed, e.g. through new information on the underlying environmental problem. In such a situation we tend to think that probably it was not such a bad thing after all that the policy objectives were not reached and intervention theories could also have a function. They could be the tools by which inconsistencies, inadequacies, or unfortunate consequences could be made explicit and examined.

## 6. EFFECTIVENESS EVALUATION AND THE GOAL-ACHIEVEMENT MODEL OF EVALUATION

Is the evaluation of effectiveness just a new name for the oldest evaluation model, i.e. the "goal-achievement model"? The rationale behind this model is simple; it is based on the question "are the results in line with the goals?" followed by the question "are the results due to the evaluand?" (Scriven 1991, 178). Despite the fact that the goal-achievement model, especially in its simplest form, has some strength, its obvious weaknesses have been the basis for the development of other evaluation models. The most evident problems with the goal-achievement model are that it disregards side effects and unanticipated effects; it does not consider costs; and the relevance of the goals is not questioned (Vedung 1997, 43ff).

All the problems concerning the goal-achievement model of evaluation also concern strict effectiveness evaluations. There is, however, a huge difference, while during the golden age of the goal-achievement model it was seen as equal to evaluation, evaluation of effectiveness is not viewed as the only form of evaluation. This makes a huge difference since it makes it much easier to expand the perspective to a multiple criteria evaluation, or to accept evaluation methods where the criterion is not at the centre, such as empowerment evaluation or peer review.

The fairly common requirements of evaluations of effectiveness and efficiency or effectiveness and impacts open up the windows to multicriteria evaluations of environmental policy. Mickwitz (2000) discussed three groups of criteria: general criteria (impact, effectiveness, sustainability, flexibility and predictability); economic criteria (efficiency either as cost-benefit or cost-effectiveness); and criteria linked to the functioning of democracy (legitimacy, transparency and equity). Economic and democracy-related criteria could also be included among the general criteria, but there are reasons for considering them separately. Economic criteria are important but their use in evaluations in the environmental context involves particular problems. Criteria linked to the functioning of democracy are important for all evaluations of public policy, but if the argument that

environmental policy imposes special challenges on democracy (e.g. Lafferty and Meadowcroft 1996) is valid, then the use of these criteria in evaluations of environmental policy is especially important.

In which way an evaluation including other criteria in addition to effectiveness becomes broader can be described by examining the different stages in the "input/output" model these criteria refer to (Nagarajan and Vanheukelen 1997, 20; Hildén et al. 2002, 19). The relevance criterion links the underlying environmental problem to the objectives of the environmental policy, by asking whether the policies, as they were established years or decades ago, actually addressed the problem in the light of today's knowledge of the environmental issues. Impact assesses to what degree outcomes, including unanticipated ones, are due to the output of the policy. While some view efficiency as a criterion that links inputs and outputs, e.g. how many permits does an environment authority with a certain number of employees and financial resources produce, efficiency is mainly a criterion linking outcomes at any stage to inputs. Sustainability and flexibility both concern the ability of the final outcomes to meet long-term needs. Transparency and equity refer to all individual stages as well as the whole system, while legitimacy refers primarily to the whole system, although problems with legitimacy may arise at a specific stage.

## 7. CONCLUSIONS

This article has argued that undertaking an effectiveness evaluation involves three steps: to identify the goals of the policy, to find the effects of the policy and to combine these two. While the three steps may appear easy, they all involve many challenges. Some of these challenges can be met by utilising some specific evaluation approaches, such as intervention theories or triangulation based on different methods and data. Other challenges, for example on whose goals the evaluation can be based, are mainly choices that have to be made based on an open deliberation of the interests of different stakeholders. Then there are some challenges that mostly affect the type of result one can expect from an effectiveness evaluation; if no quantitative goal has ever been set it is often more reasonable to produce qualitative conclusions

based on a wide-ranging discussion of the vague goals than a quantitative result based on goals taken out of the blue.

Intervention theories may have a role for all three steps of the effectiveness evaluation approach discussed, but what is the overall role of intervention theories? Intervention theories are no panacea for the problems inherent in an evaluation. The reconstruction of intervention theories is never enough; since they only deal with suppositions not the empirical world, other tools will be needed as well. Methods to collect and analyse data on outputs, outcomes, causal relationships, etc will also be needed. Evaluation designs concerning for example participation and use will have to be made. Although employing intervention theories does not per se solve these issues it is seldom an obstacle and often can make the decisions easier.

Evaluating the effectiveness of policies, including environmental policies, is very important although it means many challenges. To assess the effectiveness of an environmental policy is hardly ever enough - a broader perspective is required in order to be able to arrive at conclusions on the merit, worth and value of a policy. By the same token it should be emphasised that individual studies might sometimes focus on only effectiveness as long as a broader picture is kept in mind. If goals set by elected political bodies are not followed up, this can hardly be in line with a transparent democratic process. Broad evaluations ignoring effectiveness would therefore clearly miss a very important element. The challenges involved in evaluating effectiveness are largely there in any type of evaluation. While we cannot simply pretend these challenges do not exist or meet them with simple tools, a versatile and transparent approach is the best way to produce knowledge relevant for policymaking.

## NOTES

<sup>1</sup> A draft of this essay was presented at the 5th Conference of the European Evaluation Society in Seville, October 10th - 12th 2002; the discussions at this conference were very useful for the final essay. I would especially like to thank Professor Evert Vedung for the comments he provided prior to and during the conference. An anonymous referee of Hallinon Tutkimus provided several constructive comments to the submitted manuscript that helped in further improving the text.

<sup>2</sup> The same imprecise use of the term also occurs in Finnish. In Finnish vaikuttavuus means effectiveness and vaikutus means effects, so although clearly different from the English words the terms are still close to each other.

<sup>3</sup> Although most authors use the term "program theory" (e.g. Chen 1990, Rogers et al. 2000) the term "intervention theory" will be used (e.g. Vedung 1997, 301). The justification is that "intervention theory" is a more general term including theories of programmes, policies and policy instruments. Another issue related to the term was raised during the discussions in Seville, that is whether the word theory should be used or not. At the Finnish Environment Institute we have for some time been using the term "intervention theory" only within the evaluation community. We use terms like "assumed causal mechanisms" or "intended mechanisms" when we communicate with other stakeholders.

<sup>4</sup> As Angela Browne and Aaron Wildavsky (1983, 196) put it "Difficulties with certain studies arise when stakeholders propose competing policies while they value similar outcomes or when conflicting stakeholders propose similar policies while valuing contrasting outcomes. Politicians often seek agreement on a general policy, postponing what the parties to the bargain think they are going to get out of it or how."

<sup>5</sup> The origin of the word "scenario" is in the area of performance theatre, where it is used for the sequential elements of a screenplay, e.g. the actions of the actors or changes in the stage setting. During World War II the word "scenario" was used in war strategic analyses. It has since then been used more extensively and now occurs in many different planning settings. (EEA 2001a)

## REFERENCES

- Bartlett, Robert (1994). 'Evaluating Environmental Policy Success and Failure', in Vig N. and Kraft M (eds.) *Environmental Policy in the 1990s: Towards a New Agenda*, 2nd edition, Washington D.C. Congressional Quarterly Press, p. 167 - 197.
- Browne, Angela, Wildavsky, Aaron (1984). What Should Evaluation Mean to Implementation?, in J. Pressman and A. Wildavsky 1984, *Implementation, Third Edition, Expanded*, University Of California Press: Berkeley, 1984, 181 – 205.
- Chen, Huey-Tsyh (1990). *Theory-driven Evaluations*. Newbury Park: Sage.
- EEA (2001a). *Scenarios as tools for international environmental assessment*, Environmental issue report, No 24, Copenhagen: European Environment Agency.
- EEA (2001b). *Reporting on environmental measures: Are we being effective?*, Environmental issue report, No 25, Copenhagen: European Environment Agency.
- EEA (2002). *Environmental signals 2002 – Benchmarking the millennium*, Environmental assessment report, No 9, Copenhagen: European Environment Agency.
- EPA (1997). *The Benefits and Costs of the Clean Air Act, 1970 to 1990*, Prepared for U.S. Congress by U.S. Environmental Protection Agency. <http://yosemite.epa.gov/EE/epa/ria.nsf/vwRef/A.97.20?OpenDocument> (29.10.2002)
- Hildén Mikael, Lepola, Jukka, Mickwitz, Per, Mulders, Aard, Palosaari, Marika, Similä, Jukka, Sjöblom, Stefan, Vedung, Evert (2002). *Evaluation of environmental policy instruments - a case study of the Finnish pulp and paper and chemical industries*, Monographs of the Boreal Environment Research, 21. <http://www.vyh.fi/eng/orginfo/publica/electro/mb21/mb21.htm> (29.10.2002).
- Jänicke, Martin, Weidner, Helmut (1995). "Successful Environmental Policy: An Introduction", in Jänicke and Weidner (eds.), *Successful Environmental Policy – A Critical Evaluation of 24 Cases*, edition sigma, Berlin: Rainer Bohn Verlag 1995, 10 - 26.
- Lafferty, William, Meadowcroft, James (eds.) (1996). *Democracy and the environment: problems and prospects*, Cheltenham: Edward Elgar.
- Majone, Giandomenico, Wildavsky, Aaron (1984). *Implementation as Evolution*, in J. Pressman and A. Wildavsky, *Implementation, Third Edition, Expanded*, University of California Press: Berkeley, 163 – 180.
- Mickwitz, Per (2000). *Evaluating Environmental Policy Instruments*, presented at the 4th EES Conference in Lausanne October 12th - 14th 2000, 20 p.
- Määttä, Kalle (2000). *Energiaveropoliitikka (Energy tax policies)*, Helsinki: Kauppakaari.
- Nagarajan, Nigel, Vanheukelen, Marc (2002). *Evaluating EU Expenditure Programmes: A Guide – Ex post and Intermediate Evaluation, XIX/02 - Budgetary overview and evaluation*, Directorate-General XIX 1997, [http://europa.eu.int/comm/budget/evaluation/guide/guide00\\_en.htm](http://europa.eu.int/comm/budget/evaluation/guide/guide00_en.htm). (29.10.2002).
- Ojala, Olli (1997). *Suomen ympäristösuojelun kehittyminen (The development of environmental protection in Finland)*, *Ympäristö ja Terveys-Lehti*, 1997 (28) 3-4, 72 - 85.
- Patton, Michael (1997). *Utilization-Focused Evaluation..* Thousand Oaks: Sage.
- Rogers, Patricia, Hacsí, Timothy, Petrosino, Anthony, Huebner, Tracy (eds.) (2000). *Program Theory in Evaluation: Challenges and Opportunities*, *New Directions for Evaluation*, 87 (Fall), San Francisco: Jossey-Bass.
- Scriven, Michael (1991). *Evaluation Thesaurus*, fourth edition. Newbury Park: Sage.
- Swedish EPA (2002). *Fifteen Environmental objectives – are we on track?*, Swedish Environmental Protection Agency, Falköping.
- Vedung, Evert (1997). *Public Policy and Program Evaluation*, Transaction Publishers: New Brunswick.
- Weiss, Carol (2000). "Which Links in Which Theories Shall We Evaluate?", in Rogers et al. (eds.) 2000, 35 -45.