



**PUBLIC GOVERNANCE AND TERRITORIAL DEVELOPMENT DIRECTORATE
PUBLIC GOVERNANCE COMMITTEE**

GOV/PGC/RD(2004)4
Unclassified

**THE ROLE OF EVALUATIONS IN POLITICAL AND ADMINISTRATIVE LEARNING AND THE
ROLE OF LEARNING IN EVALUATION PRAXIS**

by Jan-Eric Furubo, Swedish Audit Office

**29th Session of the Public Governance Committee
15-16 April 2004
International Energy Agency, Paris**

This paper originally appeared in the OECD Journal on Budgeting, Volume 3, Number 3. It is presented to the Committee as additional background for the discussion on evaluation.

For further information, please contact Andrea Uhrhammer, Secretary of the Public Governance Committee, Tel. +33-1-45.24.78.32, email andrea.uhrhammer@oecd.org.

JT00161326

Document complet disponible sur OLIS dans son format d'origine
Complete document available on OLIS in its original format

**THE ROLE OF EVALUATIONS IN POLITICAL AND ADMINISTRATIVE LEARNING AND
THE ROLE OF LEARNING IN EVALUATION PRAXIS**

**BY
JAN-ERIC FURUBO***

The writing of this paper, is in reference to my contribution in *Can Governments Learn?*, one of the books published by the International Evaluation Research Group. My immediate and first reaction was that it is of course always very nice to be asked to contribute for a distinguished audience. My second reaction was to ask myself when my essay in *Can Governments Learn?* was really written and what questions it dealt with; and perhaps even more, what was the answer to the question which constitutes the title of the book. The need to ask the latter questions had of course something to do with the answer to the first question, namely that the essay was written about 10 years ago.

It was therefore obvious that the discussion in this book in a more technical sense had to be updated, but also had to take into account both some new empirical material and the discussion within a couple of intellectual fields.

The title of this current paper implies several questions: to what degree politicians and administrators learn from evaluations, and also to what degree learning takes place among the evaluators themselves. These questions create a sort of strategic problem. Am I going to give the answer at the beginning of my paper or perhaps later? But perhaps I will compromise: I will start with a short version of the answer now, but elaborate on the answer later.

The brief answer, given a Swedish perspective, is: no, evaluations do not seem to matter very much when it comes to more politically significant changes in policies. It seems that only in exceptional cases do evaluations appear to play a part in significant changes in the orientation of policies. However, in relation to the ongoing implementation of a given programme the situation is different. It seems here that evaluations are able to play a substantial role, at least sometimes. And regarding the producers of evaluations themselves, it seems that they are not learning so very much. They tend to carry out the same evaluations with the same questions year after year.

For some this can seem a bit controversial, and I guess that for others it might seem like a confirmation of what they have long suspected. However, I assume that both groups expect some elaboration on these statements and also some explanations. So I will continue by talking about three things. First, I will say something about why this question of political and administrative learning in relation to evaluations is more important today than it was 10 years ago. Second, I will give a more elaborate answer to the question about what we actually know about the use of and learning from evaluations. Third, and this is the most lengthy part, I will discuss why the decision-makers so seldom use

* Jan-Eric Furubo, who at the time when this article was written was the Head of the Secretariat for Strategic Analysis within the Swedish National Audit Office, has worked with questions related to building of evaluation capacity and evaluation strategies since the 1980s. He has published several articles and publications in the field of evaluations and budgetary questions. He was co-editor of *The International Atlas of Evaluation* published in 2002 and has served as Secretary General and board member of the European Evaluation Society.

evaluative information in relation to significant changes in policies, and why it seems different when it comes to fine-tuning and the implementation of policies.

The fundamental question which was addressed in *Can Governments Learn?* in 1994 and of course also in the chapter about the Swedish experience was: under what circumstances do governments learn through evaluations? This question is of course even more relevant today, almost 10 years later. The reason is that the diffusion of evaluation praxis and culture has taken place to a great extent after the publication of the book.

The *International Atlas of Evaluation* (“*The Atlas*”), one of the very few studies about the international development of evaluation, highlights the relevance of this question. One of the main conclusions in this book partly contradicts earlier discussions in this field. The earlier perception was that the evaluative praxis at a national level, since the 1970s, spread to more and more countries starting in a handful of pioneer countries which adopted such praxis as early as the 1960s; in other words more and more countries adopted an evaluative praxis in each period of time (paraphrase Rogers, 1995, page 23).

The Atlas shows that quite a different scenario seems to be more likely. For a rather long time – perhaps a couple of decades – no evaluative culture or praxis had developed beyond more than a handful of western countries. Then in the 1990s, a sizeable group of countries entered the evaluation era. The main difference between these two groups of countries, if we simplify things a bit, has something to do with internal and external forces. The first group, the handful of early adopters so to say, brought about an evaluative praxis due to internal forces that created a pressure and a need for conducting and using evaluations. Those countries, which adopted an evaluation praxis in the 1990s had been forced to do so, as a result of external pressure from institutions such as the European Union, the World Bank and to some extent certainly, also the OECD.

However, I am not going to discuss the implications of this difference nor what consequences it will produce. The important thing is that these external forces have led to a diffusion of evaluation praxis to more countries than ever. So the questions about the use of evaluations and political and administrative learning are certainly even more relevant today than they were 10 years ago.

Do evaluations lead to learning among:

- politicians?
- administrators?
- and under what circumstances?

These questions can be discussed in different intellectual contexts. A great deal of effort has been put into studies and discussion about different forms of utilisation and the influence of evaluations. At the same time, and partly in other intellectual circles we find discussions about organisational learning, and more specifically about learning in political environments - governmental or political learning.

But when we discuss these questions we soon find that we cannot do it at a general level. We all know that the political and administrative systems at the national level produce many different forms of decisions, and we can guess that the discussion of our questions is not quite the same when it comes to a decision about a fundamental change in policy or a minor fine-tuning manoeuvre. I will therefore distinguish between three kinds of decisions:

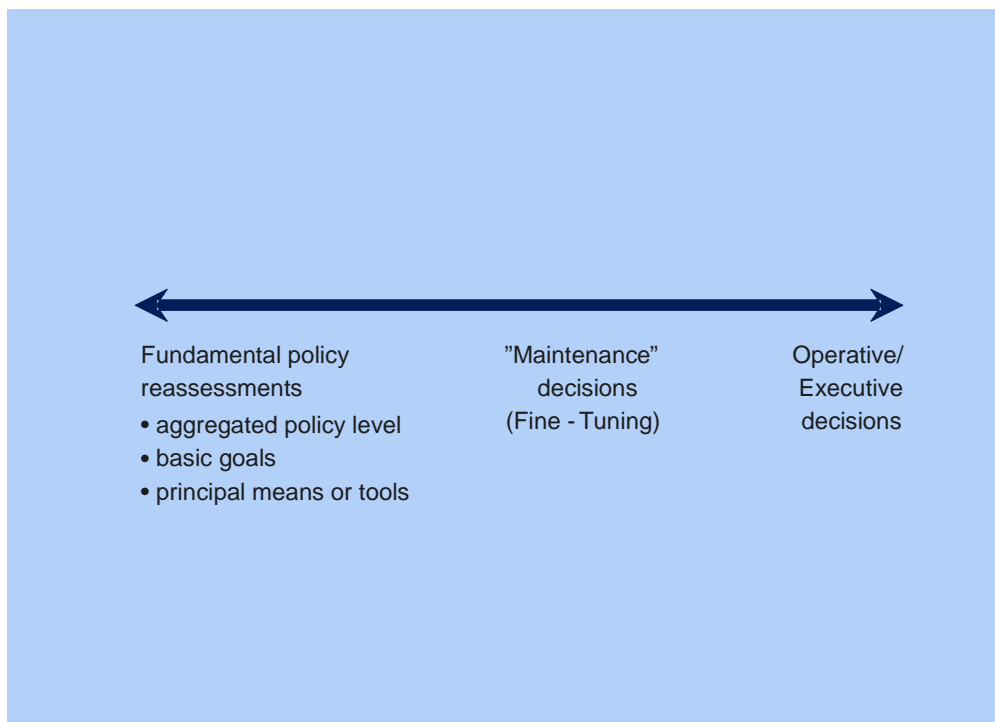
- fundamental policy reassessments;

- middle range decisions or “maintenance decisions”;
- operative decisions.

There are undoubtedly other terms, but there are also weaknesses in categorisations like this. How we place different decisions in such a categorisation is of course always a bit arbitrary, and what is placed in one category one day is put in another category next week or next year.

Perhaps it is also more realistic to talk about a scale which, at one end, has more fundamental policy reassessments at a rather aggregated policy level - decisions which can question the mere existence of a governmental policy, its basic goals and its principal means. At the opposite end of the scale, there is pure and simple technique, *e.g.* operative decisions which are part of the ongoing implementation of an intervention or a programme.

Figure 1. A scale of decisions



Between these points we can find a lot of other types of decisions, which perhaps can be labelled as a sort of policy maintenance. These decisions are related to more disaggregated goals and the use of instruments at a lower level in the end hierarchy of means available. Even if these kinds of categorisations are relative and unstable, I think they can in any case be fruitful for our discussion.

Another thing can be said about these kinds of categorisations which may be relevant for the use, non-use and perhaps the misuse of evaluations: we often meet different players involved in the decision-making process depending upon the category to which the decision belongs. In a country like Sweden it can even be said, at least on a text-book level, that the division between the ministerial structure and the different agencies, numbering about 300, reflects these categorisations. Somewhere, a borderline has been drawn between politics and implementation. On one side of the borderline we have what is regarded as implementation – the task of the agencies – and on the other side, what is regarded as policy or politics.

In Sweden it can also be said that there is a very elaborate structure for the production of evaluative information in relation to different kinds of decisions. Internal evaluations are to some extent the task of all agencies. They have to report back on an ongoing basis about their activities, and this creates a stream of evaluative information. Some agencies also contribute with special evaluations. Several bodies, research institutes, governmental commissions and so on produce evaluations of an *ad hoc* character. This structure of producers of evaluative information generates of course a lot of different information, and it is also obvious, as I will explain further, that the different kinds of decisions have to be fed with different forms of evaluative information.¹

However, even if this kind of categorisation presents some difficulties, the real problems will of course appear when we continue to the more empirical questions about what we actually know about the role of evaluations in different kinds of decisions.

1. Few empirical studies

The reason that the empirical questions are difficult is of course that we have very few empirical studies in Sweden about the use of evaluations and governmental learning in relation to evaluations. So, to a great extent, what can actually be said in this field has a hypothetical character.

This is certainly the situation at the endpoint of our scale, where we are going to start, namely fundamental policy reassessments. Studies about how different forms of evaluations have influenced decisions and changes in the central goals, their relative weight, and the choice of policy instruments are rare. What is indicated in the studies, is that evaluations have only influenced fundamental reassessments of policies in just a few, exceptional situations.

2. Evaluations do not lead to significant changes of policy...

In the chapter about the Swedish experience in *Can Governments Learn?* (mentioned above), the discussion in this respect is based on a couple of different sources.² The conclusion in *Can Governments*

¹. In this context the terms evaluation and evaluative information are used with great openness. However, even with such elasticity I do not include all kinds of descriptive information. We can assume that some notion of reality lies behind every policy or every governmental programme. The foundation of an intervention is the perception something needs to be done, in other words that a problem is occurring or, perhaps, will occur in the future. The purpose of an intervention is to reduce the magnitude of this problem or to avoid it. And an important kind of information therefore relates to the fulfilment of political goals. Securing such information is not always a simple task. But basically national statistics, as in many other countries, can indicate to what extent the “big goals” have been fulfilled. So a stream of such descriptive information is produced (Sjöström, 2002). To what extent this stream reaches the political decision-makers and in what way this kind of information is used in the political decision-making are questions which I will not include in my discussion.

But when I speak about evaluations which can be used in the reassessments of policies, I am oriented toward the idea that evaluations will give us some information about how a certain action, on a micro- or macro-level, can be judged: to what extent has the action caused a certain development and how can this be explained?

This perspective lies very near the definition by Evert Vedung (1997, page 3): “Evaluation = df. careful retrospective assessment of the merit, worth, and value of administration, output, and out-come of government interventions, which is intended to play a role in future, practical action situations.”

². The main sources are a study published by the Swedish National Audit Office (Riksrevisionsverket, 1991) in which 100 evaluations were examined and a dissertation about the Swedish commission system. In the

Learn? is that evaluations do not lead to a questioning of the basic assumptions underlying a certain policy “if we consider central policy-making processes at the level of the government and Parliament” (Furubo, 1994, page 59).³ It is further said that:

A conclusion that may be drawn from our examination of the extent to which evaluations lead to learning is that only in exceptional cases do evaluations appear to play a part in significant changes in orientation of policy.

We cannot find more than a couple of later Swedish studies which address the question of use in such terms that they are relevant in this context and they give us the same picture. In a report which is based on interviews from a special investigation group within the Ministry of Finance in September 2002, the situation in relation to what I have called fundamental reassessments is summarised in the following way: “The experience demonstrates that reassessments often have been caused by special events and usually depend on the disclosure of a special and an unsatisfactory state of affairs” (Finansdepartementet, 2002, page 56). So the limited sources at our disposal indicate that the conclusion from 1994 is valid. It is also important to note that the conclusion is not contradicted by studies of policy formation and policy shifts in different areas. In such studies related to areas such as housing policy, energy policy, crime prevention and so on, information from evaluations is not very often referred to as a major explanation for changes of any significance. The information acquired from evaluations does not seem to be a major explanation for significant policy changes.

3. But they are used in fine-tuning and implementation

Moving to the other end of the scale, to the more operative decision-making, there is some comfort in the fact that things seem quite different. When looking at this, we can find the following conclusion in *Can Governments Learn?* It is said there that “on a more technical level the situation is quite different. In such a context evaluations are able to play an appreciable role – and much evidence would indeed seem to indicate that they do.”

As indicated in the quotation, we found evidence which supported the statement about the use of evaluations in these more administrative processes, 10 years ago. Today we have of course even more up-to-date material at our disposal.

later the author studied, among other things, the effect of accumulation and knowledge within the framework of the various governmental commissions on political positions, as well as other issues (Johansson, 1992).

³. One of the key concepts of the book is single-loop versus double-loop learning. Simplified it can be said that the effects of double-loop learning on decision-making means that you question the basic assumptions underlying what you are doing. Both single-loop and double-loop learning can take place at different levels. When double-loop learning is discussed in *Can Governments Learn?* in relation to Parliament and government, it implies what is discussed here as fundamental policy reassessments. This is illustrated by the following example (page 61): “The Swedish Customs and Excises Department combats the illegal supply of narcotics in order to control the consumption of narcotics. These questions are raised at government and Parliament level as if limiting the physical supply of narcotics is a good way of reducing the consumption of narcotics. If the governments and Parliaments were to alter the activities of customs service on the basis of studies demonstrating that changes in the system of legal sanctions would provide a better way of affecting the supply of illegal drugs, then this would be an instance of double-loop learning. But we might also speak of double-loop learning if the Swedish Customs Agency had, for instance, chosen to employ dogs as a central aspect of border control in the area of narcotics, only to change this policy if it were to be shown that the use of dogs was predicated on quite erroneous assumptions with regard to the behaviour of smugglers (or dogs).”

It is obvious that the visibility of evaluative information related to more ongoing decision-making has increased. In a 1998 study, the Swedish National Audit Office compared the appearance of result information in the budget bill in 14 different areas at two points in time. The result was a marked increase in the evaluative information. Another study about the dialogue between the ministries and the agencies also indicated that the information the agencies delivered was used in the ministries' discussions about the agencies future activities (RRV, 1999).

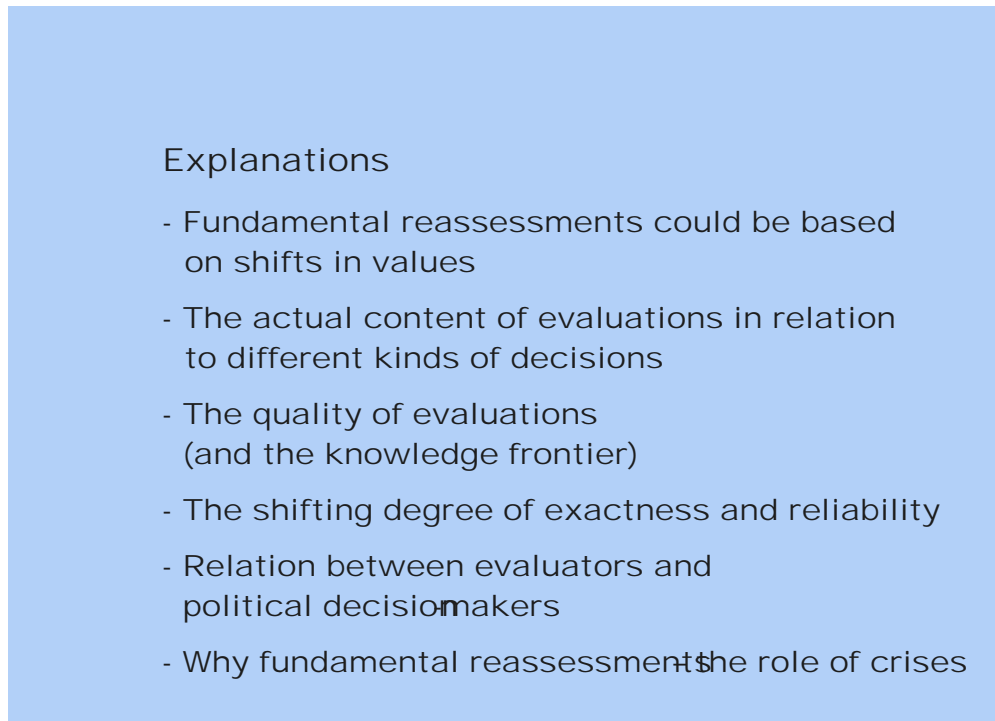
A study from 2002 (Statskontoret, 2002) indicates a further increase in relation to the extent to which evaluations are mentioned in the yearly budget bill, and the same picture is given in other reports and studies.

I therefore think it is safe to say that we have had an increase in the evaluative information which has reached decision-makers during the last 10 years. It is evaluative information related to what has been accomplished by the agencies, its quality, its cost and the internal administrative efficiency of these agencies. The orientation is more towards output and performance rather than effects and preconditions for effects.

The 2002 study shows that about one-third of the evaluations which were mentioned in the budget bill were used as a justification for present policies. About 15% of the evaluations are used as an argument for change, and in 17% of the cases, the government indicates that it will look further into the matter and then return with proposals to the Parliament. So, if we regard the cases in which the evaluations are used to justify or legitimate present programmes, more than two-thirds of the evaluations have been used.

So the different roles evaluations play in different parts of the scale is striking. How can this be explained? Can this picture actually represent the truth? Can we find reasonable explanations for this lack of utilisation of evaluative findings in relation to more fundamental reassessments; and for the fact that we so seldom use evaluations in relation to significant changes in policies? And can we explain the different degree of utilisation in relation to decisions about the more detailed construction of a policy or its implementation? Yes, in my mind it is possible to find at least some explanations that make this picture at least plausible.

Figure 2. Explanations



4. Evaluative information lacks relevance

The first explanation is that the very nature of some decisions makes the evaluative information highly irrelevant. A policy or intervention can be discussed and questioned both from a value perspective, and for more instrumental reasons. The value component can be expected to be more salient when we move towards the fundamental reassessments endpoint in the scale, and the role of empirical information increases when we move in the other direction.

So it is easy to point out that the existence or basic orientation of a programme sometimes can be regarded purely as a question of values. We can on one level imagine that all politicians agree – which is certainly not always the case – about the actual situation in some area, *e.g.* the standard of housing, the social distribution of education and so on. However, there can be quite different opinions about how a given situation should be judged. Some politicians may regard the situation as unsatisfactory, justifying a political intervention. Other politicians may regard the same situation as quite satisfactory (Sandahl, 1986).

And even in those cases in which all politicians agree that a given situation is unsatisfactory, there can still be divergent opinions concerning whether or not it is a political issue at all, *i.e.* if the situation motivates political intervention.

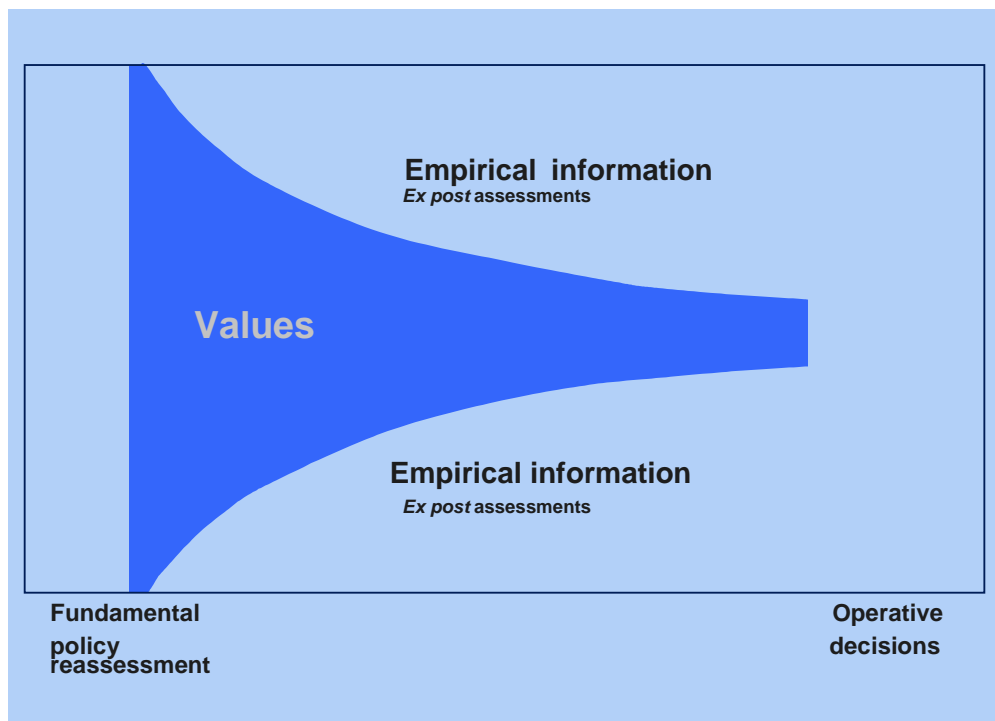
And a politician arguing against an existing governmental intervention or programme from such a value perspective has little use of information in evaluations. If the evaluations show that the existing programme has been successful, and perhaps indicates that some changes could make it even more successful, it is still irrelevant information for the politicians who are against the intervention as such. And the same can be said if the evaluations show the opposite: the intervention was a failure and perhaps quite

different instruments are needed to make the intervention successful. Even this information is of course quite irrelevant, if the politician is against the intervention as such.⁴

In these situations, information from evaluations lacks relevance and the political position towards the intervention or programme is based on values about what is good and bad, what is better and worse and what is the role of government. The position does not depend on information concerning whether or not a certain intervention was a suitable means of reaching a certain goal or solving a certain problem in society.

But a very brief moment of reflection will probably also lead us to the conclusion that decisions which concern the basic orientation of policies are value-based to a much higher extent than decisions at lower policy levels. In other words, the relative weight of values and empirical information is quite different in more fundamental policy reassessments compared with more technical and instrumental decisions concerning the more detailed construction of a given policy or its implementation.

Figure 3. The role of values and empirical information in relation to different decisions



But even if we, so to say, sort out these situations when a policy or an intervention is questioned from the perspective of values and when evaluations therefore lack relevance, politicians can still question the basic orientation of a policy or a programme for more instrumental reasons. The decisions about the future fate of the intervention could be based on an assessment of how effective the intervention is in the fulfilment of certain goals. We are then faced with the question of whether or not the tools available to the

⁴.

A similar situation is when it is agreed that something is a problem, and also that this problem is something which the government has to deal with, but some politicians do not think that they can afford it given other needs. Of course it can be said, "Yes, this is a very unsatisfactory situation and in my opinion this is something which should be resolved by governmental intervention, but we cannot afford it in light of other needs." And the conclusion may be that a governmental programme is abolished or a decision is made not to start a new programme on grounds which at least limit the relevance of information in evaluations.

government and Parliament can influence a certain development in society. And we can question which policy instruments are the most suitable, and also their basic construction in the specific situation.

We can certainly imagine that evaluations could contribute information in this kind of reassessments. So we are still confronted with the question of why evaluations do not seem to have a greater influence on fundamental reassessments of existing policies.

One part of this discussion concerns the actual character of the evaluation which are produced in Sweden and probably in many other countries. Most of the studies which are labelled evaluations in Sweden do not provide information at the fundamental goals-oriented level. Just to give an example: governmental agencies, governmental commissions and different research institutions in Sweden have produced literally hundreds of evaluations about the use of information as a policy instrument in areas such as energy consumption, health and so on. These evaluations provide a lot of information about the dissemination of the information, the efficiency of different channels compared with each other, the changes in knowledge and attitudes in different target groups which can be attributed to the information and so on. In other words, the evaluations provide a lot of knowledge about the implementation of, in this case, a certain policy instrument. However, they do not provide answers to questions concerning in which situations information is a suitable policy instrument in relation to other policy instruments, in which situations information should be used instead of other policy instruments and in which situations information could cause the opposite effect of what was intended.

There is even a tendency to produce fewer evaluations, which deal with the basic assumptions about how a certain intervention can influence different causal relationships and the circumstances for governmental interventions. There has been a tendency over the past 10 years or so to produce evaluations in a much shorter time and to provide more “easily captured” information.⁵ More and more of the evaluations are oriented toward implementation, output and performance and are part of the system of agencies reporting to the government. The information in the evaluations therefore is more relevant in decisions which are part of the maintenance of policies or operative decisions.

This explanation is empirical in character. It says that to some extent we produce the “wrong” evaluations if the intention is for them to be used in policy reassessments, and that we to some extent produce the “right” evaluations if the intention is that they be used in more operative decisions. And this is certainly not a new answer in the Swedish discussion. In more than a few governmental documents produced during the last few decades – literally – there has been a complaint that the existing evaluations do not give us help in questioning policies nor in deciding what resources should be allocated for different interventions and which interventions have been the most successful.

5. Quality of evaluations and the knowledge frontier

A closely related explanation is that many evaluations are still of poor quality, even from a more technical point of view (ESO, 1996; Statskontoret, 1999) and this is of course a problem in relation to all kinds of decisions. The report from 1996 is most harsh in its assessment of the quality of evaluations from 22 governmental agencies, evaluations conducted in relation to the agencies’ own activities.

But a more fundamental quality problem has to do with the fact that many evaluations are repeated without any clear link to earlier evaluations. For example: in the field of energy conservation, in which more than 100 evaluations were produced in the 1980s, very much the same evaluations were carried out in

⁵. Indications of this are discussed in Furubo and Sandahl, 2002 (page 126) and in several reports from the Swedish Financial Management Authority (ESV, 1999) and also in the discussion about the system of Governmental Commissions (ESO, 1998).

the 1990s. The “new evaluations” have raised the same evaluative questions in relation to the same policy instruments and with the same research design without any discussion of earlier results. And I feel rather certain that this is the case in many areas, as in my earlier example concerning the use of information as a policy instrument.

So, the evaluation process itself enjoys only a very limited degree of learning. In examining the enormous number of “day-to-day” evaluations that are made within the various administrations and research institutions, it is evident how little use is made of previous evaluations (= cloning of evaluations).

This kind of evaluative amnesia makes the movement of the knowledge or evaluative frontier very slow. Part of this amnesia is also that many evaluations do not make use of more general knowledge and rather seldom relate themselves to a more general body of knowledge, which may be highly relevant. So to some extent at least the experienced politicians and bureaucrats have understood the rules of the game. Do not trust the evaluators too much!

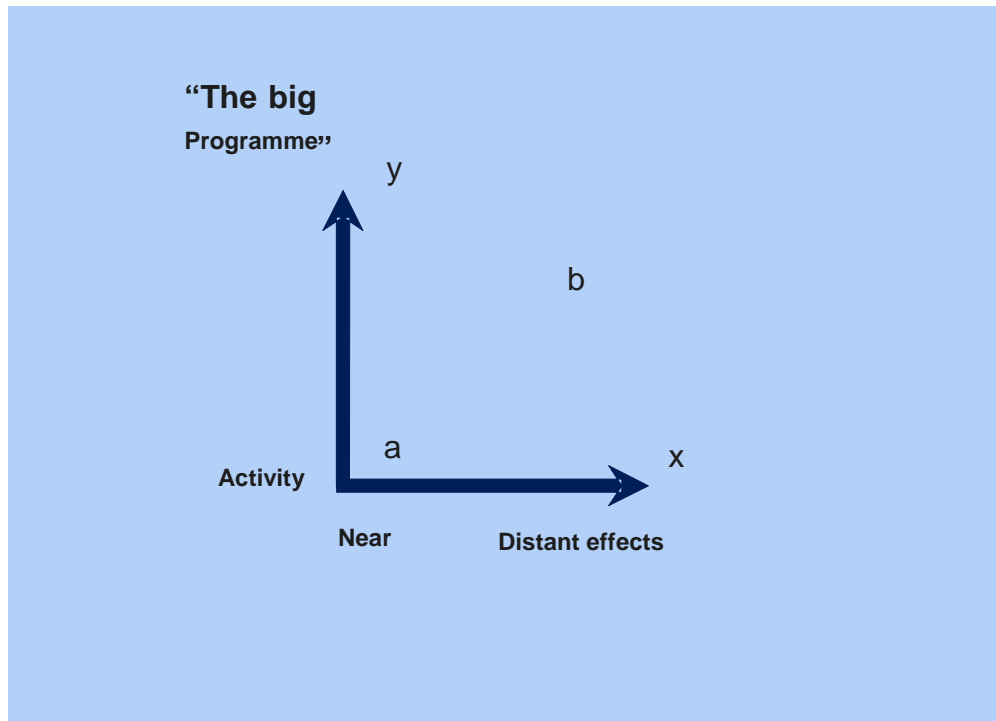
6. The shifting degree of exactness and reliability

These latter explanations have to do with the actual character of the evaluations. If we could find the main explanations following these lines, we could certainly reduce this to a question of how we could change the character of the evaluations which are produced.

But I am afraid that besides this empirically oriented answer, we also have to discuss our basic ideas about how evaluations can contribute to reassessments of policies, policy orientation and political or governmental learning or whatever we want to call it.

I have said that the role of empirical information varies depending on where we are on the scale, and it is also obvious that we need different forms of evaluative information depending on what kind of decisions we are discussing. When we are talking about how we can improve the technical construction of a given policy instrument we need a certain type of information, but when we are discussing whether or not it is a good idea to use this policy instrument at all, we need of course quite different types of information. It is obvious that these different forms of information have many different characteristics. The first type of information is often of a character which makes it suitable for the ongoing production of information, which perhaps is not the case when we talk about the latter type of information.

Figure 4. Different evaluative statements



And in this context, it is relevant to make a distinction about how exact and reliable different evaluative statements are. This is illustrated in Figure 4. On the Y-axis we can imagine, nearest the origin of co-ordinates, a limited measure, *e.g.* an information brochure on influencing household use of energy. A little further up we find the “package of information efforts” required to have an effect on energy consumption, and still higher up the total measures needed to influence energy use in society. On the X-axis we can imagine some kind of chain of effects where, in the example given, we see the reception of the information nearest the zero, and farthest to the right the influence on the environment and the national economy that reduced energy consumption would have.

It is easily seen that we are talking about completely different kinds of information in “a” and “b” and, without crossing over the line to a discussion about theories of knowledge, we – and certainly the decision-makers – can expect information in “a” to be very exact and, at the same time, reliable. Evaluative statements on the effects of complex interventions in still more complex social processes and courses of events are of a quite different nature. An evaluation that draws conclusions about the extent to which a reform of the school system aiming to influence equality between men and women has actually changed vocational choice, relationships in the home, etc., can hardly do so particularly well, and in any case not with a high degree of reliability. The nearer to “b”, the more improbable it is that different evaluations will come to the same conclusions.

The problem is of course that the “b” information is more relevant than the “a” information, when it comes to decisions about the basic orientation of a policy. It is therefore difficult to imagine that this kind of often uncertain evaluative information should be transferred in a more immediate way to the political decision-making system. The lack of use in relation to fundamental policy reassessments can therefore seem very rational from the politicians’ point of view. But on the other hand, we can also expect that the decision-makers make more use of the evaluative information in relation to the maintenance of policies and operative decisions. The information they need in these processes is more of the “a” character and

therefore also more reliable. The difference when it comes to the use of evaluative information in different kinds of decision-making processes seems therefore reasonable even from this perspective.

7. Relation between evaluators and political decision-makers

This brings us to the question of the relationship between the evaluative camp and the political camp. Discussions, even the more recent ones, about the use or influence of evaluations, have had the evaluation and perhaps the evaluation process as starting points. The information in the evaluation has to be disseminated for the purpose of leading to different forms of use. On this point, it can be added that the discussion about the different forms of use has long since passed the stage at which the perception of use was an immediate response among the decision-makers to a specific evaluation.

But even if we are all aware of the sophistication of this discussion about use and utilisation, the underlying idea is still very much based on a model which starts with the evaluation, which hopefully will be disseminated to the decision-makers, who will react to the findings in one way or another – directly or indirectly.⁶

This model or notion of the utilisation process is perhaps a realistic one when it comes to implementation on a more administrative or operative level. But in reference to the relationship between evaluative information and political decision-making, this notion is perhaps too simplified.

To what extent the knowledge in different evaluations can influence policy reassessments probably also has something – and perhaps a lot – to do with the existence and the character of intellectual structures which are feeding the political system with knowledge. This factor is discussed in a study about the Swedish Stabilisation Policy 1975-95. The author (Jonung, 1999) tries to “explain the sequence of policy switches that characterise Swedish Stabilisation Policy during the period 1975-95”. The author highlights the importance of intellectual structures from which the political decision-makers may possibly receive information. He points out the existence of a profession in close contact with the political system by saying, “politicians responsible for policy also obtained information, inspiration and arguments from the economic profession, *i.e.* from economists active at universities and research institutes”. Further he points out the role of international organisations as another source of information.

If we relate the result in this study to our discussion, it would give rise to the question about the need for such intermediate structures between the evaluation side and the political side: intellectual structures which digest the results of evaluations of – usually – rather limited programmes or activities and relate them to a more general body of knowledge.

In the case of the Swedish Stabilisation Policy, it was easy for the author to point out a well-defined academic discipline or profession with channels to the political elites. Perhaps we can imagine that the role of evaluations in political decision-making varies depending on the existence of such intellectual or professional structures.

So instead of the notion of a model in which evaluations are regarded as knowledge channelled directly to the political decision-makers, an alternative way of looking at things is to use the analogy of a “knowledge bank”. Different evaluations can contribute to this knowledge bank with information deposits. The officials of the bank, to continue the metaphor, interpret the information delivered in these different

⁶. Vedung (1997, page 265ff) discusses several forms of use with reference to (among others) Weiss (1972) in terms of Response Steps in a Communication Theory based on McGuire (1989). Weiss (1998, page 305ff) discusses use and dissemination in relation to decision-makers and other stakeholders. Owen and Rogers (1999, page 105ff) are strongly oriented toward dissemination to decision-makers.

studies, rearrange the information and relate it to earlier knowledge in the field. They finally communicate the knowledge in their possession to the political elites.

That is to say: the immediate users of evaluations are not the decision-makers, but the officials of this metaphorical bank. So the extent to which the information which has been gained in connection with earlier governmental interventions will actually be channelled into the political and administrative system very much depends on the character of these intermediate knowledge structures.

The existence of such knowledge structures is of course very different in different areas. Sometimes there are well-defined knowledge structures coinciding with academic disciplines. However, this is often not the case, and this makes things a bit more difficult. Further, the lack of such structures can perhaps to some extent explain why evaluations are used as little as they are.

But what I want to point out is that the discussion about how evaluations can contribute to more fundamental policy reassessments perhaps should not focus on the relationship between evaluations/evaluators on one side and the political system on the other. Another perspective focuses on a more triangular relationship, and asks questions about the relationship between the evaluation community and other knowledge structures on the one hand, and these structures and the political system on the other hand. It includes questions about the existence of different channels between these more general knowledge bodies and the political system, and also to what extent evaluations can contribute to this more general knowledge.

To strike a more optimistic tone, one of the lessons learned in Sweden is related to the question about structures to channel more general knowledge into the political processes. In *Can Governments Learn?*, I described to some extent the system of governmental commissions, which since the beginning of the 20th century has been a way to channel scientific knowledge into the political system. And in the rare, exceptional cases in which evaluations seem to have influenced decisions about policy orientation, they seem to have not done so directly, but through such more information-digesting structures.

8. The role of crises

Finally, I intend to add one more possible explanation. When it comes to implementation and decision-making concerning how a certain agency should conduct its tasks, a system can be created more or less forces the decision-makers to react – at least on some level – to the stream of evaluative information and take it into account in making decisions.

When it comes to fundamental policy reassessments, the situation is different. Since the 1960s, Sweden has tried, on several occasions, to incorporate fundamental reassessments of the basic commitments of the state and the basic tools which should be used to reach different goals in more bureaucratic systems. These systems have aimed to guarantee that such reassessments are conducted regularly, every third or fifth year, for example. In short, and with some brutality, it can be said that these efforts have failed.

The reason is probably not because problems connected with the implementation of these systems, different budget reforms and systems for the production of both output oriented information and outcome oriented information. Instead, and more probable, it is the underlying assumption regarding what motivates that kind of fundamental reassessment of the commitments of the political institutions, and the central means for the fulfilment of these commitments.

These assumptions have probably been too naive. The political level does not start to question a fundamental policy just because one or two or even five or six evaluations point out different problems, *e.g.* the lack of goal fulfilment and questions about the underlying rationale for the policy.

Instead, in the discussion about policy shifts, the importance of crises or extreme situations is often stressed. And if we follow this thought it leads us to several questions about how, and in which situations, it is fruitful to bring in information from evaluations in different decision-making processes, and it also give a background for the previous discussion about the role of different knowledge bodies which can provide the answers when the politicians really want them.

9. Finally

Perhaps one or two of the things that I have said can be deemed a bit controversial and if so I am glad. I admit I have tried to say a couple of things which certainly could be disputed. I have tried to moderate expectations of the role of evaluations and, in doing so, I have found it important to bring to light a couple of distinctions. They can seem too simple, but even so I think that they are important.

We often talk about evaluations in a more general way, thereby hiding the fact that the information provided in evaluations varies in a number of respects. And at the same the decisions in the political and administrative system are of many different kinds.

These basic facts have consequences with regard to what role we can expect evaluations and evaluative information to play. They also have consequences for the discussion about how frequently different forms of information should be produced and for questions about responsibilities and the organisation of the commissioning and conducting of evaluations as well as for the relationship between evaluation and other forms of knowledge production.

And we can probably imagine several different developments in this field. On one hand, we have perhaps a tendency towards a more ongoing, continuous stream of evaluative information, which perhaps will be channelled more or less directly to the administrative decision-makers from different systems. This kind of information is generated in more or less direct relation to the implementation of different activities. It gives the agencies, or more generally speaking the organisations responsible for the implementation, a key role in the production of such information.

On the other hand, we have the development of other forms of evaluations which have to be done on an *ad hoc* basis, and which much more strongly than today need to interact with a more general science production. So we have quite an agenda for our future discussions about the role of evaluations in political and administrative learning and also about the role of learning in evaluations.

BIBLIOGRAPHY

- ESO (expertgruppen för studier i offentlig ekonomi, the Expert Group on Public Finance) (1996),
Kan myndigheter utvärdera sig själva? (only in Swedish).
- ESO (expertgruppen för studier i offentlig ekonomi, the Expert Group on Public Finance) (1998),
Kommittéerna och bofinken. Kan en kommitté se ut hur som helst? (only in Swedish) DS 1998:57.
- ESV (Ekonomistyrningsverket, The Swedish Financial Management Authority) (1998),
Resultatinformation i budgetpropositionen - Före och efter budgetlagen (only in Swedish), 1998:46.
- ESV (Ekonomistyrningsverket, The Swedish Financial Management Authority) (1999),
Informella kontakter i samband med regleringsbrev och årsredovisningar (only in Swedish),
1999:19.
- Finansdepartementet (Ministry of Finance) (2002),
Regeringskansliets kontroll och styrning av statlig verksamhet (only in Swedish).
- Furubo, J.E. (1994),
“Learning from Evaluations: The Swedish Experience,” in F. Leeuw, R. Rist, and R. Sonnichsen
(eds.), *Can Governments Learn? Comparative Perspectives on Evaluation and Organizational
Learning*, New Brunswick, NJ: Transaction Publishers.
- Furubo, J.E. and R. Sandahl (2002a),
“Introduction – A Diffusion Perspective on Global Developments in Evaluation” in J. E. Furubo,
R. Rist, and R. Sandahl (eds.), *International Atlas of Evaluation*, New Brunswick, NJ: Transaction
Publishers.
- Furubo, J.E. and R. Sandahl (2002b),
“Coordinated Pluralism – The Swedish Case” in J. E. Furubo, R. Rist, and R. Sandahl (eds.),
International Atlas of Evaluation, New Brunswick, NJ: Transaction Publishers.
- Johansson, J. (1992),
Det statliga Kommittéväsendet – Kunskap, Kontroll, Konsensus (only in Swedish – English
summary), Edsbruk, Sweden: Akademitryck AB.
- Jonung, L. (1999),
Med backspegeln som kompass – om stabiliseringspolitik som läroprocess, published by ESO
(expertgruppen för studier i offentlig ekonomi, the Expert Group on Public Finance, only in Swedish
– English summary), DS 1999:9.
- McGuire, W.J. (1989),
“Theoretical Foundations of Campaigns” in R.E. Rice and C.K. Atkin (eds), *Public Communication
Campaigns*, Newbury Park, California: Sage.

- Owen, J.M. and P.J. Rogers (1990),
Programme Evaluation – Forms and Approaches, London: Sage Publications.
- Rogers, E.M. (1995),
Diffusion of Innovations (Fourth Edition), New York: Free Press.
- RRV (Riksrevisionsverket, The Swedish National Audit Office) (1991),
Att mäta resultatanalysen – Vem analyserar vad, hur mycket och på vilket sätt (only in Swedish).
- RRV (Riksrevisionsverket, The Swedish National Audit Office) (1998),
Resultatinformationen I budgetprocessen – före och efter budgetlagen (only in Swedish).
- Sandahl, R. (1986),
Offentlig styrning – en fråga om alternativ (only in Swedish), Stockholm: Riksrevisionsverket (The Swedish National Audit Office).
- Sjöström, O. (2002),
Svensk statistikhistoria (only in Swedish), Södertälje: Gidlunds förlag.
- Statskontoret (Agency for Administrative Development) (1999),
Mittutvärderingar av strukturfonderna – en övergripande utvärdering, 1999:53.
- Statskontoret (Agency for Administrative Development) (2002),
Utvärderingar – Av vem och till vad (only in Swedish), 2002:21.
- Vedung, E. (1997),
Public Policy and Programme Evaluation, New Brunswick, NJ: Transaction Publishers.
- Weiss, C.H. (1972),
Evaluation Research: Methods for Assessing Programme Effectiveness, NJ: Prentice Hall.
- Weiss, C.H. (1998),
Evaluation (second edition), Upper Saddle River, NJ: Prentice Hall.