

Finding out about results from projects and programmes concerning democratic governance and human rights

A study commissioned by Sida

Written by Kim Forss

September 2002



SWEDISH INTERNATIONAL DEVELOPMENT
COOPERATION AGENCY

Department for Democracy
and Social Development
Division for Democratic Governance

Table of Contents

Executive summary	3
Chapter 1. Introduction	6
Background	6
Purpose	7
Methods	7
Guide to the reader	8
Chapter 2. The tools for accountability	8
Conceptual clarity	8
What is accountability?	9
The missing link – results	9
Performance management	11
Evaluation; process and product	11
Performance rating	13
Indicate – to state briefly	14
Concluding remarks	15
Chapter 3. Assessment of results in development cooperation	15
The quest for accountability	16
Evaluation or performance management, or both?	17
Organising an evaluation system	18
Some notes on systems design	19
The quest for indicators	20
Concluding remarks	20
Chapter 4. How do you explain impact?	21
The practice of finding results	21
Attribution and causality	22
Non-linearity and complex systems	23
Structuring the inquiry with models	24
The methods of investigation	24
Measurement and scales	25
Quality of evaluation	26
Concluding remarks	28
Chapter 5. How do we evaluate democracy “at home”?	28
Can democracy be evaluated?	28
Evaluating democratic institutions	29
Evaluating governance – civil service reform	30
Evaluating human rights	31
Concluding remarks	31

Chapter 6. Conclusions and recommendations	32
Common frameworks?	32
Models for accountability	32
Models for evaluation systems	33
Models in evaluation	33
Approaches to accounting for results	33
Approaches to indicators	34
Approaches to objectives	34
The road ahead	35
References	36
Annex 1. Terms of reference	38
Annex 2. Annotated bibliography	43

Executive summary

The purpose of this study is threefold; (1) to present a review of the “research frontier” in respect of methods to assess results of projects in democratic governance; (2) to assess what other development cooperation agencies do in this field, and to suggest lessons to be learned from their best practice; and (3) to discuss specific methodological issues on evaluation, performance management, rating systems, etc.

In Sweden as well as in other OECD countries, governments are concerned about the strength of democratic traditions, and there are many projects that aim to engage the citizens in public affairs. There are many evaluations of the state of democracy as such, as well as of democratic institutions and governance. Evaluations tend to be interdisciplinary and are characterised by methodological pluralism. Many evaluations take a participatory approach, as they aim to reinforce public participation in politics, while at the same time assessing situations. They have a practical focus on how to improve governance, democracy and human rights. Even if evaluations are theoretically and conceptually thorough, they seldom imply causal connections between project activities and overriding objectives.

The starting point is to analyse changes and seek explanations, rather than to analyse project performance and seek effects. It is generally understood that evaluations do not provide final answers, they enlighten the debate, guide decision-making and extend knowledge. It would seem as if evaluations of domestic projects and programmes are given far larger resources than is generally the case in development cooperation, but still the demands on what kind of results that can be proven are more modest – or realistic is perhaps a better word.

Evaluation, performance management, results, rating systems and indicators are all words that need to be defined. The concepts overlap and administrative practice is built on poor definitions – or no definitions at all. This paper argues for a distinction where evaluation is the most encompassing term; evaluation is the systematic inquiry into the worth or merit of an object. When you evaluate, you assess results on several levels, you may also use some kind of rating system, and indicators may be used, for purposes of rating or as part of a description. But an evaluation is more than an assessment of results, or a performance measure. Evaluations seek explanations, and account for why and how things happen, and also arrive at value statements.

Performance management builds on rating systems that are subjective and quantitative, but some rating systems are built on objective measures. Rating systems are built on indicators, which are weighted. An indicator can be subjective or objective, quantitative or qualitative. When organisations present results, they usually do so in quantitative terms, and it is not considered reliable to base results on subjective indicators. Still, there are several organisations in development cooperation that base their results on subjective and quantitative indicators.

Accountability in development cooperation has over the years followed many tracks. In the past, it was considered most important that donor agencies had evaluation systems in place that, if necessary and called for, could present results to the public and to the political authorities. In recent years, there has been a gradual shift to emphasise accountability through measures of indicators. So far, not many agencies have developed systems that can provide reliable and valid indicators of impact. The main problem is that indicators only capture a small part of reality. They are useful means to “anchor” and develop a qualitative analysis, but on their own, they are meaningless if not directly misleading.

Donor agencies take different approaches to presentation of results, and there are no common models employed in the design of evaluation systems. However, when it comes to the practice of writing terms

of reference, commissioning studies, and using the results, there are many similarities. The handbooks and manuals that are developed for the professional staff can often be shared. But when it comes to the practical task of actually doing an evaluation, there are as many approaches (or models) as there are evaluations. This is to be expected, as evaluation is usually a preliminary set of conclusions about events. Both evaluators and those who commission evaluations look for novelty, and strive to improve over previous standards.

Evaluators (in developments assistance, as well as elsewhere) have a basic toolkit of surveys, interviews, observation and document analysis to work with. However, these methods can be used with tremendous variety, and there many technical issues to be observed in the design of instruments, practical issues relating to the choice of method, and finally methodological choices, that is, relating to the scientific foundations of the inquiry. The evaluation process must be a congruent whole of techniques/instruments, methods and methodology.

Approaches to evaluation are built on assumptions of causality and systems complexity. This report distinguishes different types of causality. It contrasts a mechanistic and reductionist approach to the subject of evaluation, to an understanding of social phenomena as complex, dynamic and non-linear systems. Evaluation methodology has to deal with cluster effects and qualitative leaps in non-linear systems. Social sciences have made a great deal of progress with the development of chaos theory and the sciences of complexity. With inspiration from these fields, it is possible to address problems in a more realistic, more relevant, and hence also more useful way.

It is always a challenge for an evaluator is to structure the inquiry. The most important tool that can be used is a model of the phenomena to be studied. A good descriptive model of the relationships between events is must. With the help of one or several models, it is possible to portray and assess quite complex phenomena. It is still relatively rare that evaluators in development cooperation make much use of models. However, if one is to make sense of complex causal chains this is necessary.

A key message in this study is that it is always possible to evaluate, that is, to systematically assess the worth or merit of an object. It is not necessary that projects or programmes are planned in certain ways, for example with the help of logical frameworks. Nor is it necessary that goals or objectives comply with some specific standard of clarity, brevity or precision. There is a certain stage in the evaluation process where clear objectives can be of help, but a thorough evaluation usually has to go beyond the obvious statement of goals and objectives to assess whether these were relevant and realistic. There can be other reasons to strive for clear and precise objectives, but it is not required for the sake of evaluation.

So, what are the main suggestions to be derived from this study concerning how to account for results in the field of democratic governance and human rights:

1. Put the emphasis on evaluation research, which is the instrument that allows the organisation to handle complex and multidimensional phenomena.
2. Develop the quality of evaluations by emphasising the need for modelling the research context, structuring the inquiry with the help of relevant hypothesis, and encourage methodological pluralism.
3. When allocating resources, consider that it is better to do some few evaluations of high quality than many with questionable validity and reliability.
4. Learn from domestic experience that proper evaluation research takes time and money, if it is to be useful.

5. Develop the communication with the audiences who are concerned with results, and address their needs for short, relevant, and easily digestible information based on evaluation findings.
6. Performance management systems can be applicable when accounting for results at lower levels, mainly relating to expenditure items and descriptions of activities, but they are not appropriate for analysis of results at the level of goals and purposes.
7. Indicators are tools that can be used at several levels: (1) to outline development trends, (2) to provide inputs to the qualitative analysis, and (3) to describe results at the level of activities. There is a need to develop the methodological competence to work with indicators at each level, according to the role that can be expected of the instrument at that level.

Chapter 1. Introduction

Background

Several events over the past decades have led those working with development cooperation to reconsider the concept of development, in particular how systems of governance and politics relate to growth, economic development and poverty alleviation. The boom and subsequent depression in East Asia, and the collapse of the planning economies in Eastern Europe, are only two well-known examples of institutional changes that have an impact on democracy, governance and human rights. In fact, a theory of development, which does not adequately cover the nature and quality of governance and political systems would today appear meaningless.

But governance is a complex subject. Even though significant research has been devoted to it over the past decade, the state of the art is anything but coherent. Many of the actors most concerned with global development have had problems bridging the gap between theory and practice. One part of the problem is manifest in the practice of evaluation; the methods are at best uncertain, and the practice ranges from the excellent to the miserable. There is a need to take stock of the state of the art, and to ascertain that evaluation systems provide information so that actors can learn from experience and be held accountable for what they do in respect of the development of democracy and human rights.

Progress in democracy and governance are closely linked to the concept of development itself. In fact, taking a broad look at the development literature, it seems as if good governance (in the widest sense) not only explains development, but also is synonymous to development. When practitioners and policy makers in development cooperation nowadays turn to governance, they are going back to the core issues of the reasons for aid.

In Swedish development cooperation, democracy and human rights could always be seen as an aspect of the overriding objective to “further the living conditions of the poor” (*främja de fattiga folkens levnadsvillkor*), which was formulated in the policy document 1962:100. In the beginning of the 1990s, democratic governance and human rights moved to the centre stage as a primary objective of development cooperation. There has been a real increase in the number of programmes and projects that have explicit objectives within these spheres.

However it is one thing to start up and implement programs in respect of democracy and human rights, but yet another thing to account for results. Evaluation is never easy, but seems to be particularly difficult with respect to changes in governance, political regimes, or the respect for human rights. The difficulties are not due to any shortcomings among development agencies, but rather lie in the intellectual heritage of evaluation itself. Evaluation is a cross-disciplinary subject, with roots in political science as well as in public administration, business and management, organisation theory and scientific method. Evaluation has focused on the assessment of results in economic terms. Early on, cost-benefit analysis and similar methods of assessment came to dominate the discipline. There is no doubt a tradition of applying evaluation for purposes of control of expenditure programs, where the inquiry has been more concerned with questions of “value for money”, than other possible aspects of worth and merit.

Purpose

The Government of Sweden¹ has requested Sida to develop the methods to assess results in respect of the development cooperation objective “to promote democratic governance”. The request was formulated in 2001, and during that year the first steps were taken through a number of seminars to deepen the competence base at the section responsible for programmes in this field. During 2002 the work is continued, for example through this study. The purpose of this study is threefold²:

1. To present a review of the “research frontier” in respect of methods to assess results of projects in democratic governance;
2. To assess what other development cooperation agencies do in this field, and to suggest lessons to be learned from their best practice;
3. To discuss specific methodological issues on evaluation, performance management, rating systems, etc.

With this information, it is expected that the division of Sida (DESA)³ will acquire a common approach to the possibilities, as well as the problems, in the evaluation of democratic governance. A common understanding is expected to contribute to a sound and relevant application of evaluation methods to assess results, learn from experience and improve projects and programs, and also to a clearer and more substantive dialogue with partners, and more stringent reports to the government on impact and effectiveness.

Methods

This is a desk study. There is thus no particular methodological challenge in the conduct of the study itself. The problem is to find information, and for that I have relied on:

1. Literature searches in the OECD database
2. Searches in the scientific journals in the evaluation field (Evaluation, Evaluation Review, Evaluation and Program Planning) over the past 5 years
3. Net searches on the websites of major aid agencies, bilateral as well multilateral
4. Searches among publishers of scientific texts in evaluation and political science
5. Personal requests for information from the Swedish government, from the OECD working group on public management, and from the Swedish system of public inquiries
6. Reviews of major scientific studies on the subject of democratic governance.

It is obvious that I have not found all information there is to be had. Every new search yields more information, more interesting studies, new projects in methodological development. I'm afraid there is no way to guarantee that I have found the best examples of good practice, nor even all relevant major studies in the field. But I do hope the information on the following pages will be sufficiently rich to stimulate the

¹ Sida's regleringsbrev 2001 och 2002

² Terms of reference are enclosed in annex 1.

³ Sida's Department for Democracy and Social Development (avdelningen för Demokrati och Social Utveckling – DESO) is responsible for most of the activities that relate to democracy and human rights, and within it the Division for Democratic Governance (Enheten för Demokratisk Samhällsstyrning) is the substantive unit, which develops projects and programmes. The activities of DESA cover democratisation, human rights, good governance and public participation.

thinking among DESA personnel, and possibly others who come into contact with the paper. I have listed relevant search engines, web resources, publications, where those who are inclined can continue where I break the search. The present study was undertaken during six weeks of July and August 2002.

Guide to the reader

The vocabulary is rather confusing in the field of evaluation, accountability and performance management. Chapter 2 is devoted to the issue of definitions. The chapter suggests definitions, analyses how terms are used and could be developed, and how they fit in a wider system of accounting for results in development cooperation.

Chapter 3 turns to the policies and practices on how to account for results. It points to the history of aid evaluation, and the inherent tensions in the objectives of an evaluation system. Many of the technical terms used in evaluation are brought up, and so are different approaches to accountability. The chapter illustrates the debate around indicators and performance assessment, and ends with a review of evaluation quality.

Chapter 4 is devoted to specific methodological issues, particularly on how to treat causality, how to explain impact, and how to work practically with models and hypotheses. This chapter aims to clarify the choices that have to be made when confronted with a practical evaluation assignment, and it suggests ways and means to strengthen the quality of evaluation.

All professions have their “blind eyes”, and in development cooperation one blind spot is that we often forget the experiences from our own countries. There is a tendency to invent the wheel anew. Thus chapter 5 discusses the domestic experience, that is, looking at how we evaluate progress in democracy and human rights in Sweden and in other OECD countries.

The annotated bibliography provides some examples of books, papers, websites, etc. on the subject of assessment of democratic governance and human rights.

Chapter 2. The tools for accountability

Conceptual clarity

The overriding question that is treated in this paper is how an organisation can be held accountable for its activities. How can government, parliament, and ultimately the general public, be certain that the funds allocated to promote democratic governance in development cooperation really are well spent. This is what accountability is all about.

An organisation can use a number of administrative devices to create accountability; there are monitoring and evaluation systems, performance management systems, rating schemes, indicators of performance, and the like. The problem is that these terms can be understood in different ways. They are related to each other, but it is not quite clear how.

The fact is that there are no commonly agreed definitions. Practice varies between organisations, and the systems that work well in one context (with the use of a particular set of definitions) do not work well in another setting. However, it can be very difficult to communicate about these subjects if one

does not speak the same language. Hence the purpose of this chapter is to set out a framework of terms that can be used in the following sections of this paper.

What is accountability?

As the overriding purpose of this paper is to contribute to “accountability”, it might be useful to start with a fuller understanding of that concept. Accountability is defined as the “responsibility for the justification of expenditures, decisions, or the results of one’s own efforts”. Project managers should be, it is often said, accountable for their project’s achievement. Accountability thus often requires some kind of explanation of how one spent the money (fiscal accountability), but furthermore it is expected that one is able to justify this in terms of achieved results.

Let us take an example. Teachers have sometimes been held accountable for their students’ achievement scores, which is of course entirely inappropriate because their contribution to these scores is only one of several factors (the most frequently cited other influences are student ability, support from parents, support from peers, and support from the rest of the school environment outside the classroom). On the other hand, a teacher can appropriately be held accountable for the difference between the learning gains in his or her pupils and those of other teachers of essentially similar pupils.

A common fallacy associated with accountability is to suppose that a fair assessment of results requires the formulation of precise goals and objectives if there is to be any accountability. In fact one may be held accountable for what one does within even the most general conception of professional work (Scriven, 1991). Less specificity makes valid measurement more difficult, but not impossible. Captains of ships are held accountable for their actions in wholly unforeseen circumstances. But a measurement process has to be very carefully selected and applied. It would often be the case that a range of alternatives would have to be recognised by the testing process or that the process would look only at rather general features of what was done.

Ultimately it is the actor who seeks accountability who has to say whether a given measurement is to be accepted or not, to let it be known whether he, she or it, is satisfied with the results presented – in whatever terms these are presented.

The missing link – results

The assumption in this paper is that accountability is best achieved when an organisation presents results⁴ that correspond to the expectations and standards set by those who hold the organisation accountable. But what kind of results are we speaking about? The term “results” is one of those words that seem to be precise, but which actually can mean many different things. Collins Dictionary defines result as “something that ensues from an action, policy, etc.; outcome, consequence.” Some organisations focus on results in terms of profits, other look for results in terms of market share or service delivery.

Development cooperation was at one point accused of defining results mainly in terms of spending money. However that may be, there has for long been a focus on the delivery of projects, and of producing immediate benefits to those target groups associated with projects. When we now speak of results based management, it is usually implied that we look for results in terms of impact on the broader social, economic and political development in partner countries. However, the assumption needs to be tested. At the same time, it is nowadays common to turn to tools of performance management that are more applicable for “lower” levels of results, that is, results in terms of service delivery, speed and ex-

⁴ The terms of reference focus on “resultatredovisning”, which, loosely translated into English means accounting for results. It is more administratively correct to associate “resultatredovisning” with a term such as “results-based management”, or possibly with management by objectives, where the focus then is on how the achievement of objectives is measured.

tent of implementation, etc. Indicators and rating system are well-known and appropriate instruments to assess results in such terms.

A result from an activity could, in theory, be described with reference to any standard of assessments. Sida's evaluation manual indicates five "dimensions" that could form the basis for whether the results of a programme are satisfactory or not. These are efficiency, effectiveness, impact, sustainability and relevance. However, it is important to note that neither one nor all of these terms together form a result. They are characteristics of results; results should be efficiently and effectively delivered, have an impact, be sustainable and relevant. But what is it actually; what is the difference between the result and its impact?

One possible approach is to assume that results should be expressed in direct connection to the formulation of objectives; that is, the word results can only be meaningful when it is used to assess whether objectives were achieved. If the objective of a project was to strengthen the capacity of an organisation, then the result is that capacity was strengthened (or the lack of results, that the project failed to strengthen capacities).

It follows from such a definition that results can be assessed at many different levels. A project or a programme that is financed by Sida has usually been subjected to a logical framework analysis. This means that goals, purpose, outputs and activities have been specified⁵. In theory, it would be possible to describe results at each and all of these levels. Results can be described in terms of delivering inputs, producing outputs, and so on, up to reaching a development objective (goal). The problem is that the further up one comes in the hierarchy, the more difficult is it to document the results.

When discussing results within the field of democratic governance, the partners in the discussion have to define at what level they want to identify results. Nobody else can do that for them. In theory, results can be described in relation to the goals expressed by the Government, or at the level of projects and programmes. The main point to note here is that the starting point must always be an objective, and the result consists in whether that objective was reached or not. Table 2.1 below describes levels of results, and points to qualities that can be desirable in respect of each level of results.

Table 2.1 Assessment components and how these relate to the hierarchy of objectives*.

Results can be depicted at the level of:		Results can be assessed in terms of		
Goals		Impact	Sustainability	Relevance
Purpose		Effectiveness	Impact	Sustainability
Outputs		Effectiveness	Impact	Sustainability
Activities	Efficiency	Effectiveness		

* In practice, it may happen that objectives are formulated in terms of one or the other aspect of assessment, which may seem to make the cells in the table redundant. The problem would rather be that objectives are not formulated in terms that are independent of the assessment whether they are reached or not.

⁵ These terms are defined in Sida's evaluation manual, which is to be published in late 2002. The terms are also defined in the OECD/DAC handbook on methods and procedures in aid evaluation (OECD, 1986).

It may seem as a subtle point, but such a use of the term is a departure from the common-sense use of the word “results” in daily language, which only means that something is produced as an effect of some activity, policy or the like. In managerial and administrative terms, results has relevance when used to assess the achievement of objectives, and it can be confusing when that rather strict sense of results is used simultaneously as the everyday term result.

Performance management

The concept of performance management arose in the postindustrial world of the 1980s and blossomed in the 1990s. It incorporates key features of past efforts to reform the management of public service systems and programs. What is new is the context in which the concept has flourished. Government deficits have reduced program funding, leading to greater selectivity in what programs are to be given continued support and what new initiatives are to be launched.

Decentralisation of program authority has resulted in the need for central government levels to retrieve some level of control over devolved programs, as a trade-off for their loss of power. Furthermore, economic realities have generated new central government and public demands for evidence of program results. In the process, the definition of accountability has shifted from a previous emphasis on program processes to a more singular focus on program results.

Performance management is similar to results-based management. The Swedish term for both of these would be “mål – och resultatstyrning”, which was introduced in public administration in the early 1990s. The new management model replaced the previous standard of “management by objectives”, which was criticised because it failed to take into account the process whereby results were delivered, as well as what these results actually were.

Monitoring (or tracking) program outcomes with the use of performance measures or indicators, is one of the central features of performance management. Performance management is a planning and managerial tool and the drive to develop performance measures has supported much logical and strategic thinking at all levels of government. The major activity in designing performance management systems is the development of these performance measures. Measures can be developed to represent any type of variable that is of interest to those setting policies for the system.

A rating system is one kind of monitoring system, but not the only one. On the contrary, it is possible to consider other kinds of performance management systems. Rating has the advantage of integrating measures across conceptually different kinds of activities, and measures can be chosen to account for different contexts of the activities that are rated.

Performance management is usually seen as one possible approach to how higher levels in an organisational system can hold lower levels accountable for activities. Systems of performance management are often seen as contradictory to systems of evaluation (Blalock, 1999; Davis, 1999). In practice, evaluation can be part of a system of performance management, but it would normally only be a small part of such a system. So what is then evaluation?

Evaluation; process and product

Evaluation research is a young field, and though there are several professional organisations, annual conferences, scientific and popular journals, university programmes as well as shorter training opportunities, there is actually no common definition of the subject. The closest one can arrive at is the

definition sanctioned by the American Evaluation Association (Joint Committee, 1994), which says that:

“Evaluation is the systematic inquiry into the worth or merit of an object”

This is a brief and elegant definition, and it is also the least common denominator of what researchers in the field have been able to agree upon. There are a few things to note about the definition. First, the definition does not specify that evaluations have to be independent; that is, undertaken by independent, unbiased, experts.

Evaluations can also be internal, and they can build on stakeholder participation. Second, an evaluation is not defined by its purpose. Whether undertaken for control, learning, decision-making, or if it is undertaken to legitimise decisions, postpone decisions, or even to divert attention, the study can still be called an evaluation. Third, the definition does not specify what the object is. An evaluation can assess a project, a programme, an organisation, a policy, or even an object in the grammatical sense (such as a car, a washing machine, a paper mill).

It is remarkable that the definition does not specify what constitutes worth or merit. This could be goal achievement, efficiency, effectiveness, relevance, beauty, durability, survival capacity, or any other quality, or combination of qualities. Note also that the definition does not say anything about how the evaluation is disseminated; whether it should be formally presented, open to the public, or even if it has to be presented in writing. Presumably, an evaluation process could end in a seminar, and it does not necessarily have to lead to a written final report. However, an evaluation does have to be systematic, which means to show its methods and its data. Hence, some form of report is usually assumed, and thus the evaluation is both a process and a product.

Some organisations choose to specify more precisely what evaluation is in their context. It is, for example, common to specify that evaluations have to be undertaken by independent experts. The definition above is firm in two respects in particular; first that evaluation has to be an assessment of worth or merit. This distinguishes evaluation from research (pure and applied), which does not necessarily have to arrive at a value assessment. Second, evaluation has to be a systematic process of inquiry; meaning that it has to build on the methods of social science research. The assessment must build on a systematic collection and analysis of data.

Before the above definition was established, there was a proliferation of definitions and considerable confusion and anarchy. The OECD/DAC expert group on aid evaluation did a commendable job when it managed to make aid organisations agree on a definition, which is as follows:

An evaluation is an assessment, as systematic and objective as possible, of an ongoing or completed project, programme or policy; its design, implementation and results. The aim is to determine the relevance and fulfilment of objectives, development efficiency, effectiveness, impact and sustainability. An evaluation should provide information that is credible and useful, enabling the incorporation of lessons learned into the decision making process of both recipients and donors.”

Today, this definition constricts the development of evaluation as it is not sufficiently clear on the need for scientific rigour in value assessment, while at the same time the definition specifies a number of issues that do not necessarily need to be specified. In fact, the statement is more of a policy and an expression of intent than a definition in the scientific use of the term. In this paper, I will take for granted that evaluation is “a systematic inquiry into the worth or merit of an object” – no more and no less.

Evaluation differs from monitoring in that the latter is a management function for data collection. Monitoring is a regular and systematic collection of data that can be used to assess progress and per-

formance. Monitoring systems are usually helpful devices in an evaluation, but the monitoring activity as such does not usually extend to the kind of analysis, explanation and understanding that is at the core of the evaluative process.

Performance rating

Monitoring can be quite similar to performance rating; it is a systematic and regularly reoccurring activity, based on some few indicators that are measured. But monitoring as such does not always extend to rating on a scale, which is typical of performance rating systems. It is important to be clear about the objective of a performance rating system. Two objectives come to mind; the first is to strengthen the implementation of projects by tracking the results more effectively and thereby discover problems earlier, and trigger action to solve problems. The second objective is to use the system to present results, that is, to hold the organisations accountable for the performance indicators that are presented. Both are possible, but they do not live easily together. Systems that do not make clear choices, risk to fail on both objectives. A rating system must also be easy to use, and it should not require lengthy research processes to gather information.

Incentive and sanction systems are sometimes incorporated into rating systems to ensure compliance with performance expectations. Incentive systems may tie performance management to budgeting, or even to career planning and personnel evaluation. The risk is high that the data in the system is corrupted, and it may become impossible to use one's own organisation to obtain data. It is not uncommon that professional staff in organisations lack commitment to new rating systems and they often distrust indicators of performance. Such concerns are usually well founded, as many rating systems force staff to reduce the complexity of projects they work with.

Rating systems are not a substitute for evaluation. On the contrary, evaluation can contribute to rating exercises, and may enable people to assess situations quicker and more accurately. But where evaluations generate understanding, explore context and causal links, the rating system focuses on rapid assessment and outcomes, and is quantitative rather than qualitative.

No projects are similar, but application of a rating system does presuppose that there are some standard issues that are of relevance on all projects. If the projects are very dissimilar, the rating becomes incomparable, and thus it loses relevance. Organisations that have a very diverse project portfolio need not bother about rating systems.

Rating systems appear to have become popular shortcuts to accountability in development cooperation over the past few years. Most of the multilateral development banks have developed and introduced performance rating systems. The systems are different, but the African and Asian Development Banks apply systems that are quite similar to that of the World Bank, but the Interamerican Development Bank has introduced a rather different system. The differences are both of technical and of "philosophical" nature; as for example the issue of whether to present a single aggregate rate for each project, whether to connect the rating system to performance and career planning. Other differences concern the scales to use, how to connect to project implementation, and whether to share the results with other organisations.

Some organisations have developed systems, but have not yet introduced them (for example the OPEC Fund), others have actually decided to wait and see (French and Japanese organisations in development finance). KfW is the only bilateral agency that has developed and introduced a rating system, and which presents the results of the rating in annual reports. Among the Nordic agencies, Sida has developed a rating system, but it is not yet really launched in the organisation. Danida and Norad are in the developmental phase, and Norad is applying its rating system on pilot basis in two or three countries. The aid programme of the European Union has a rating system, and this in fact goes back in time to

the 1980s, making it the oldest system in development cooperation (it was called the traffic light system, as it built on flagging projects as green, yellow or red).

What makes a performance rating system good? Here are a few attributes that come to mind, and that could be used to compare organisational experiences of such systems:

1. Quality of the information; the value given on the respective indicators must be trustworthy
2. Utility; the rating should assess issues that have practical consequences – either you do something about it, or you know that you can allocate your time to other problems
3. Efficiency; the system should provide information at reasonable cost, and should also be developed with an eye to “value for money”.
4. Integration with instruments for project planning and monitoring
5. Ease of access to information in the system
6. The extent to which the system requires others (partners, clients) to spend time furnishing information for the system

The rating systems used by the banks fall short on several of these criteria, particularly numbers 3, 5 and 6, but to some extent also on number 1. The rating systems developed by the few bilateral agencies as well as by the European Union may fall short on 1 and 2, but also on number 3. To be true, that assessment is mine, based on observation and some comments from people associated with the system. It is quite surprising that no systematic comparative assessment of performance rating systems has ever been done, despite the significant amounts of money that have gone to such projects.

Indicate – to state briefly

Indicators are used to present results, to organise the data collection in an evaluation, to structure performance management and rating systems. Indicators are – or can be – building blocks in all the processes described above. What is an indicator all about? Indicator is the noun derived from the verb “to indicate”. To indicate means (1) to point out or to, and (2) to state briefly, show indirectly (The Merriam Webster Dictionary). In management use (Lusthaus et al, 1999), an indicator is an item of aggregated data, which easily and briefly provides a lot of information about an issue. An indicator is a measuring device that allows you to clarify and measure a concept. Indicators make a concept more tangible, give you something to measure, and allow measuring to take place over time. Let us look at some common examples that we are all familiar with:

If you buy a car, you would like to have some indicator of quality, and hence you turn to a consumer advisory service. The consumer advisory service may give you some of the following indicators of quality:

- number of accidents the car is involved in compared to other cars
- number of consumer complaints following the sale
- a statement on quality orientation from the manufacturer
- the type of brakes used and their durability
- servicing routines and guarantees.

These five, plus presumably several others will give you an indication of the quality of the car. It will not be the full truth, as that would be an extremely complex matter to go into, but it will give you an indication. In a simple situation such as this, it is likely that an ordinary consumer will use several indicators of quality – and before purchasing the car he or she will probably consider price in relation to indicators of quality, safety, beauty, fun and several other aspects that we could consider when buying a car.

If we are buying shares in a joint stock company we would also use indicators of performance. The stock market typically lists some 5 to 10 main indicators, such as dividends, earnings in relation to capital, price compared to earnings, earnings compared to profits, and profits compared to dividends. To be meaningful, each of the indicators would have to be studied as part of a time series. The direction and speed of change would normally be more interesting than the level at a certain point in time. Furthermore, the values of the indicators are mainly interesting and meaningful in comparison to other related indicators, for example of another company, or compared to any other form of investment.

Consider the first example. It is a rather simple thing to assess the quality of a car. Still, if you were prudent, you would probably want to have some information on at least some 20 to 30 indicators. Also, some of them would point in the direction of high quality, other to low quality. In the final end, you would have to weigh them together if you want to buy the car. Proceed from this very simple example to the assessment of democratic governance. It should be obvious that any use of indicators would have to be very carefully tailored, there would need to be really many indicators, and the process of weighing them together would be delicate.

Indicators can be qualitative and quantitative. It is common to associate indicators with plain figures, ratios and the like, but an indicator could also be a descriptive text. There is no absolute limit on the length of such a text, but there are those who used qualitative indicators based on several pages of description.

Concluding remarks

The rhetoric of how to hold an organisation accountable is surprisingly vague. Words that appear plain, such as results, indicators, and rating, are in fact understood differently in organisations. In this chapter five common concepts, from accountability, via results, to performance management, evaluation, rating systems and indicators were described.

There are three points to remember; (1) an organisation must be clear about how these words are used, and it is often better to apply scientific definitions than to build on everyday use of language, (2) the organisation must unite, it is no good if one section applies one definition of, for example, evaluation, and another section another definition, and (3) it is often useful to adopt definitions that are commonly used in other organisations. It makes communications and exchange of experiences easier. Clarity, unity and coordination with others are the key words.

Chapter 3. Assessment of results in development cooperation

The quest for accountability

Aid management has always used evaluation to account for its results. Indeed, accountability in development management has almost been synonymous to the widespread application of evaluation. It is through evaluation that Sida, as well as other bilateral and multilateral aid agencies, have been able to identify results.

Evaluation of development projects/programs need not differ much from evaluation in other (domestic) contexts. The basic purpose, as well as the methodological frameworks, and questions of quality,

utility, ethics, etc. are the same. There are differences, but so is there between evaluations in different fields on the domestic scene (compare an evaluation of military capacity to one of agricultural crop experimentation). Above all, wherever you evaluate, those who do the evaluations must know the substance of the work being evaluated.

Evaluation has always been applied to development cooperation, in Sweden at least since 1948. The question, then as today, is if aid really works – if it promotes development. Many evaluators have risen to the challenge of providing an answer to that question. Major scholarly studies such as Cassen and associates (1986), Riddell (1987), Mosley (1987), Berg (1992), Maren (1997) and the World Bank (1998) have assessed aid effectiveness in general – including governance, human rights, and institutional development. Studies of effects at country level, or within specific sectors are prolific, as are evaluations of specific projects and programmes.

It is commonly said that evaluations are undertaken for three specific purposes; (1) control, (2) management, and (3) learning. When development assistance is challenged, both its defendants and its critics point to evaluation findings. It is usually evaluations that document “white elephants”, corrupt practices, inefficient delivery systems and harmful side effects that fuel criticism. But it is also evaluations that point to successful projects and programmes that reach targets and improve living conditions in the host countries, and it is through evaluations that one can document good practice and learn from experience (Cracknell, 2000).

The recently completed public inquiry⁶ on development assistance confirms that there is a fluid border between research and evaluation. It points to the need for a diversified approach to evaluation, and to a close cooperation between academic institutions and the organisations for development cooperation. The evaluation policies and systems should take into account the differences between countries. Does the committee express an increased concern for evaluation and accountability? No, not as I read the report. Obviously evaluation, control and performance management are important topics, but not more so now than they were 10 or 20 years ago⁷. There is a difference in emphasis. Where the directives of 10 and 20 years ago focused on the need for evaluation, the latest public inquiry speaks more of accounting for results (resultatredovisning). The two could be the same, as evaluation usually describe worth or merit by pointing to results.

Evaluation or performance management, or both?

Performance management and evaluation have some similarities, but there are also considerable differences. The purpose of evaluation is primarily to increase our understanding of the major relationships imbedded in the design of programs and in explaining intended as well as unintended effects. Evaluations are different; some are exploratory, others descriptive, quasi-experimental, or experimental. There are goal-free evaluations⁸ as well as participatory evaluations. The evaluation process can be adapted to a wide number of circumstances and purposes. Performance management, once a system is designed, is less flexible. It builds on quantitative data, rapid assessment, and standardised definitions of results. The more simple and one-dimensional the subject, the more appropriate is performance

⁶ SOU 2001:96. En rättvisare värld utan fattigdom. Betänkande av den parlamentariska kommittén om Sveriges politik för global utveckling.

⁷ Comparing the above report to SOU 1978:61, Biståndets organisation. Betänkande av biståndspolitiska utredningen, and DsUD 1990:63, Bra Beslut: Om effektivitet och utvärdering i biståndet.

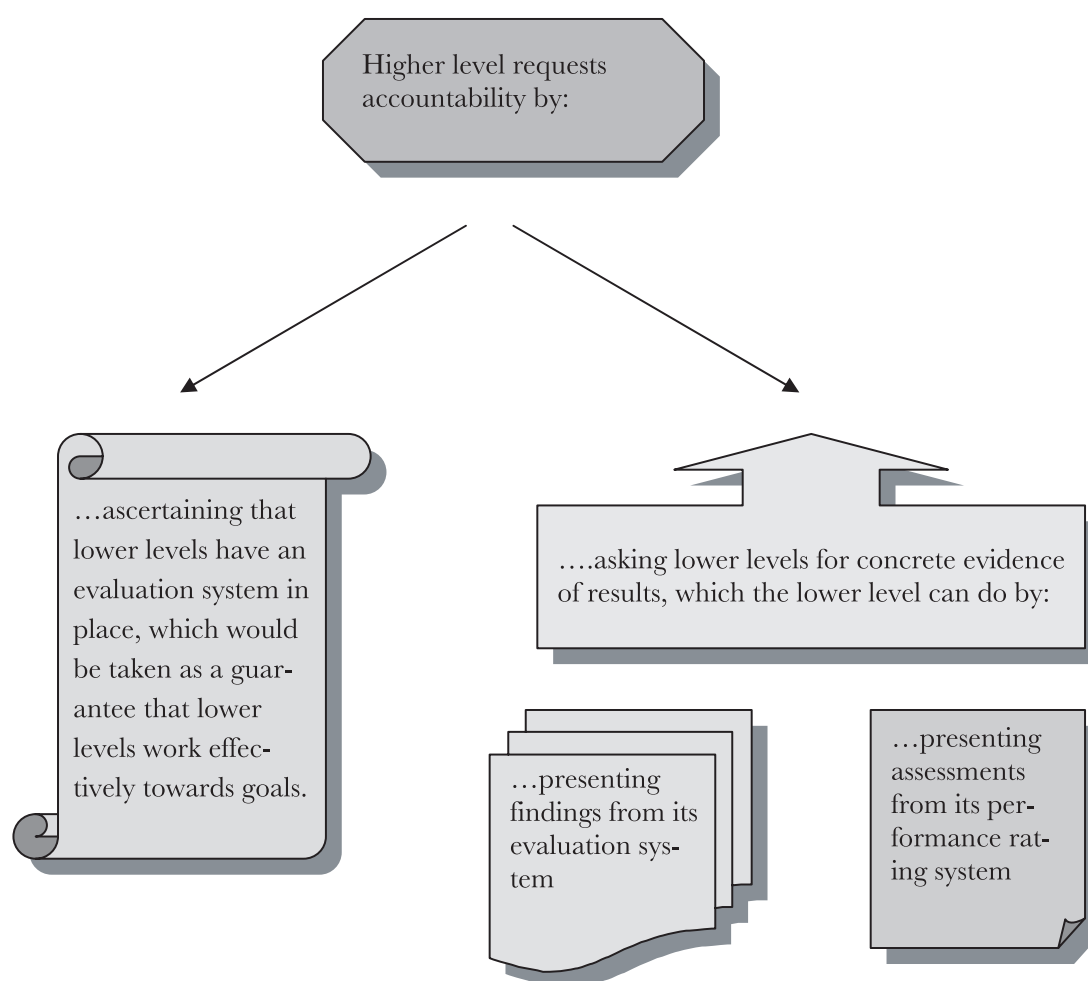
⁸ Goal free evaluation was spearheaded by Michael Scriven. The approach is based on that evaluators are not informed about the goals of the project; they are to identify what it has achieved – no more, no less. Scriven argues that knowledge about goals bias evaluators to find the expected effects, and to sidestep unintended effects.

management as a source of accountability. The more complex and multi-dimensional, the more appropriate is evaluation⁹.

There are two strategic approaches to accountability. The first approach would be to refrain from direct reports and instead make sure that the organisation itself knows what it is doing. A board could ascertain that an organisation takes the subject of organisational learning seriously; allocates adequate resources to it, invests in information technology to store and retrieve its accumulated experience, etc. So, if the organisations have a good evaluation system, there would no need to send detailed reports. This approach would be particularly relevant when goals are vague and difficult to measure, or when the activities of the organisation are very dissimilar

The second approach is to request reports back on the achievements and performance of the organisation in respect of the targets that are established. This would presumably give an accurate account of results, provided that goals and objectives are clear, and that the organisation has resources enough to provide the requested feedback. Within this approach, there are two distinct possibilities of how to account for results. The first would be to present an aggregate report that summarises the findings of an evaluation system. The second would be to use a performance management system, built on a rating scheme. Figure 3.1 illustrates the choice between the three ways of accounting for results.

Figure 3.1 Approaches to accountability.



⁹ The journal "Evaluation", number 2, volume 5, 1999, has a number of articles on the differences, similarities, advantages and disadvantages of evaluation and performance management.

The approach taken to aid evaluation has often been of the latter sort; that is, as long as the organisations have proper evaluation system, this would be a guarantee that they can be held accountable in respect of a diversity of project, country and sector objectives, but they would not necessarily have to present aggregate measures of performance. With the advent of performance management systems, the approach appears to be shifting. There is at present a tendency to ask for direct and concrete feedback in respect of aggregate targets. It is not obvious that the conditions for that kind of results-based management are present.

Which are then the factors that should determine the approach to accountability? Well, if the higher level can assess whether an evaluation system works or not, and it can “order” the lower level to put an effective evaluation system in place, this would surely be the most cost-effective way of ascertaining that objectives are reached. The two approaches in the second category assumes that the higher levels have the time and competence to take part of rather detailed reports on substantive achievements. The first of these, though aggregate findings of an evaluation system, would be suitable when programmes are very diverse, when goals are qualitative and open to interpretation. The third approach, though performance rating, would be suitable when goals are one-dimensional, when activities are standardised, and the assessment criteria simple.

Can one combine the three approaches? Yes of course, they are in fact combined most of the time. But the lack of clear choices often results in considerable confusion, uncertainty, waste of resources and low accountability.

Organising an evaluation system

As evaluation figures as important instruments in two of the approaches to accountability, let us now take a closer look at that instrument. (It should be obvious that some form of evaluation system would be the most appropriate means of accounting for results in respect of democratic governance and human rights.) Evaluation is a process that needs to be organised. The evaluation function can be centralised to a specific evaluation unit, or it can be dispersed. Speaking of development cooperation, evaluation can be seen as a function of a Ministry or it can be seen as a function to be carried out at lower administrative levels (and presumably higher levels as well, as Parliament could have a stake in aid evaluation).

Practice varies among countries and in addition the multilateral organisation have many approaches to the organisation of evaluation. In most countries, and organisations, the approaches vary over time. In Swedish development cooperation, Sida had a fragmented evaluation system during the 1970s and early 1980s (Forss, 1985), which was later formalised and centralised, and then again dispersed on several agencies and presided over by the Ministry for Foreign Affairs in the late 1980s and early 1990s (DsUD 1990:63). In 1995, it was reorganised, and now Sida has a central evaluation unit, but the line units in the organisation are responsible for evaluations of projects and programmes, while at the same time the Ministry for Foreign Affairs has a function for policy analysis and evaluation. The system has been evaluated, and there are many ideas on how to improve it¹⁰. Apart from the evaluations undertaken within the system of development cooperation, Swedish aid has also been evaluated occasionally by the National Audit Board, as well as by the Parliament’s evaluators (Riksdagens Revisorer).

Focusing attention on evaluation systems in development cooperation rather than on specific evaluation units raises complex issues. The choice of design for an evaluation unit has consequences for the

¹⁰ Nilsson, B. Et al; Utvärdering av Expertgruppen för studier av utvecklingsfrågor (EGDI) samt avseende former för en förstärkt utvärderingsverksamhet inom utvecklingssamarbetet, SINOVA, Stockholm

overall system. Evaluation may actually be going through a change process much like that of strategic planning in the corporate world in the 1970s. When strategic planning rose as a prestigious management science many corporations established centralised planning units. Two decades later, strategic planning is deemed far too important a function to be kept at central headquarter levels, instead there are many types of strategic planning diffused throughout organisations (Mintzberg, 1994).

The point that needs to be made here is that a division or a department at Sida cannot choose its approach to results without anchoring this in an organisation-wide approach to evaluation. Systems to create accountability by reporting results at one level of the organisation are either confined by, or enabled by, approaches taken at other levels. I have elsewhere argued that practically oriented, decentralised and informal systems of evaluation have the potential to satisfy several objectives at the same time (control, decision-support, and learning), while at the same time being cost-effective. However, it is beyond the scope of this paper to treat the whole evaluation system of Sida, and hence this is written against the background of the evaluation policy that is presented in the draft evaluation handbook.

Some notes on systems design

Whether working with an evaluation, or some other system of investigation, there are basically four different methods to collect data; interviews, questionnaires, observation, and study of documents (Patton, 1982). These are usually identified as the evaluation methods, but they are the methods available for any kind of investigation into results.

As applied to aid, evaluators have found themselves in the interface between academic research and public administration (Berlage and Stokke, 1992), with the needs of the administrators holding the upper hand. The most common means of evaluation is a fact-finding team. The team visits the area of the project, reports back to the agency on its findings and makes recommendations of changes. It seems simple, but several contradictions are involved. One basic contradiction can be seen in the tension between the aid administration's demand for quick answers, to be transformed into immediate action, and the professional concern of the evaluator for the methodology and precision of the craft (Thomas, et al, 1998).

Evaluation in development cooperation belongs to the more complex and methodologically challenging tasks. The questions formulated are often ambitious; evaluators are asked to look for connections between fairly minor inputs in terms of resources (projects and programmes), and lofty targets in the form of national and regional development, human well-being and institutional development. The logic of projects and programmes builds on linear connections between a cause and an effect, a simplistic image of the world and science that has been reinforced with the use of logframe techniques for planning and implementation. In reality, the context of assistance is complex and results depend on a number of other factors (Uphoff, 1992).

The process of data collection in evaluation of development cooperation is particularly sensitive. Language differences require evaluators to be very careful about how questions are interpreted and what the answers actually mean. It is useful to work with – or through – local evaluators, but these also need to be sensitised to such issues. Cultural differences may also affect the evaluation process, as for example the propensity to come forward with criticism, the perception of risks in technical choices, and the assumptions about what constitutes worth and merit (Lonner and Berry, 1986).

Evaluators in high-income countries may be confronted with equally challenging tasks, for example if evaluating the effects of a major tax reform or the development of new industrial sectors. But evaluators here usually have access to greater resources, apply diverse methods, and still have a humble ap-

proach to the complexity of the task. Evaluators need to have a sophisticated understanding of the phenomena of non-linear systems and multiple causality, and they should be able to communicate such an understanding in the presentation of evaluation results (de Vaus, 2001).

Sector programmes generally need a combination of research methods. Institutional development is an objective that needs particular attention to the choice of methods, depending on the type of institutional change analysed and the depth of analysis (Forss, 2001). Human rights and promotion of democracy are also fields that challenge evaluators. The basic elements of the evaluation process are the same, but the combination of methods, the design and application of interviews, observation, etc. must be applied with skills and with an understanding of the limits and the relativity of fact-finding in complex environments.

The quest for indicators

Many aid organisations have struggled to find global indicators and measures of performance. It has so far been futile attempts, even though lucrative for a number of consulting firms. Even if it is possible to define indicators at the level of particular projects, it is prudent to consider contextual information when decisions are taken. The more aggregate the level of analysis, the more necessary is it to supplement the data of a performance management system with qualitative and context bound knowledge. Beyond the analytical level of a sector or a thematic area indicators usually become meaningless, if not actually misleading.

In the debate on how to assess progress in democracy and human rights (as well as more generally on evaluation), there are those who favour the use of indicators, and there are those who are against them. In the latter group, one of the more vocal opponents is Thomas Carothers (1999, p. 291):

“The effort to assess the impact of democracy programs by using highly reductionist indicators is a deeply flawed undertaking that is consuming vast resources, producing little useful insight or knowledge, and introducing serious distortions into the designing and implementing of such aid.”

UNDP’s Human Development Report makes extensive use of indicators, and their methodological arguments defend that approach. The starting point is actually that there are no unambiguous, uncontroversial measures of democracy and human rights. Hence researchers have two options. They can use objective measures, such as voter turnout, the existence of competitive elections, numbers of political prisoners, etc. They can also use subjective measures, where an audience (preferably expert opinion) is asked to rank phenomenon on some scale. A classical example of such subjective measures is when business managers are asked to assess the level of corruption in different societies¹¹.

The objective measures will only cover some aspect of democracy, and so will of course the subjective measures too. The respondents only provide answers to what they are asked about. My own opinion is that indicators and indexes, such as presented by UNDP, the World Bank, Freedom House¹² and others in a development context – plus similar statistics in our own societies – never can be accepted at face value. But they are interesting and useful starting points for a better understanding of political conditions. But they must be framed in a qualitative discourse.

Concluding remarks

This chapter distinguishes three approaches to accountability. It is assumed that there is a higher or-

¹¹ Transparency International’s Corruption perception Index.

¹² UNDP’s Human development report for 2002 presents a review of different sources of objective and subjective indicators (pages 36 and 37). Other efforts, such as the Afrobarometer, World Values Survey, etc could be included in the group.

ganisational level (in our case the Ministry for Foreign Affairs and Parliament) and a lower level, which is requested to account for results (Sida). The higher level could either request the lower level to make use of an evaluation system, and seek evidence that this system works well (for example by analysing the system). Another approach is to ask for concrete evidence of results.

There are essentially two ways then for an organisation to show concrete results (provided that there are any). The first is to distil the results documented through an evaluation system. This presupposes that the organisation does have a well-working and reliable evaluation system, with clout and integrity. The second approach is to present the higher level with assessment of results based on a performance management system.

The latter approach is suitable when and if the projects and programmes are standardised and when it is easy to capture results with the help of a limited number of indicators. It is a way of reporting that does not require much of either time or other resources at the higher level. The major drawback with the former approach (aggregate evaluation findings) is that those who are the addressees (Ministry for Foreign Affairs, Parliament) have to penetrate into the complexities described in evaluation reports, and they have to understand and accept that reality is multi-dimensional, and that there are usually positive as well as negative aspects of programmes.

Indicators are necessary within the conduct of an evaluation, but quantitative indicators always need to be framed with a qualitative understanding of context and background. Simple and unqualified indicators can, in the context of accounting for results, lead to either of two mistakes; (1) they are simply neglected because it is obvious that they are not sufficiently valid to base any results on, which just means that the money spent on them was a waste, or (2) decisions are taken on the basis of unreliable and invalid data, which is likely to lead to poor decisions, which means that the system is not just useless, it is actually harmful.

Chapter 4. How do you explain impact?

The practice of finding results

This chapter turns to how the process of evaluation can be used to find, document and elaborate on results. It is of course so that an evaluation system does not “deliver accountability” unless it is a good system, and unless the evaluations as such are good.

Evaluations can describe and analyse results at all levels, in respect of activities and outputs, as well as in respect of purpose and goals. Evaluations can assess worth and merit on each of these levels in terms of efficiency, effectiveness, impact, relevance and sustainability. Furthermore, evaluations can detect undesirable effects, can assess worth or merit above and beyond goals and purposes, and can be used to reflect on and question these.

However, it may be a waste of resources to analyse activities with the help of evaluation. Evaluation is a powerful tool, but it is also costly. Some questions are better answered by more simple and thus cheaper means. Performance rating systems can, if correctly designed, be useful instruments to track progress and to assess results at lower levels, and in uncomplicated terms. However this chapter deals with the “higher levels of results” (from the explanation of impact and onwards, and with the starting point at the level of outputs).

Attribution and causality

It is quite interesting to note that most texts on the evaluation of democracy, good governance and institutional change in the context of development cooperation express despair over the question of attribution. So what is attribution? The Collins Concise Dictionary explains the verb attribute as follows: “to regard as belonging to, produced by, or resulting from, for example to attribute a painting to Picasso”. Attribution is the noun formed from the verb. So attribution is an aspect of causality, which would be the more scientific term to use when you analyse why things happen, and what effects they have.

None of the texts that I have perused on the evaluation of democracy, governance and human rights in the OECD countries speak of attribution. But in development cooperation, there is hardly a report that does not complain about the impossibility of its task because of the problems of attribution. Crawford and Kearton (2002) devote several pages to the problem, OECD/DAC (1997) outline many difficulties connected to attribution, such as:

- intangible lines of causality
- vague nature of outputs, cannot be distinguished from other donors
- project outcomes may be cancelled out if not supported by other policies and processes.

A nine-volume Danida study concludes that “identifying a wider impact of specific inputs is in most cases nearly impossible¹³. Sida commissioned a study of the evaluability of Democracy and Human rights projects¹⁴. The authors conclude that the difficulty of attribution is a serious evaluability deficiency, as it concerns the evaluator’s ability to establish causal linkages between the project and its effects.

So why is attribution a problem only in studies of development assistance, but not on the domestic scene? One reason may be that evaluation in developed countries generally has more resources, better skilled evaluators, and more time at its disposal than does evaluation in development cooperation. Another interpretation could be that the problems exist, but are discussed in other terms – and solved. Methodological sections of evaluation reports describe how they deal with causality, they do not say “here is a problem that is difficult to deal with”. A third reason may be that those who commission “domestic” evaluations have more realistic expectations of what to expect in terms of results, and they set more realistic targets for the evaluation. Yet another interpretation could be that we here have an example of a post-colonial mode of thinking, where funding agencies and “benevolent” donors treat “recipients” arrogantly and superficially, and face them with demands for evidence of results that one would never expect on the domestic scene.

However, let us turn from attribution to a discussion of causality. In the most rigorous definition, one phenomenon (X) is said to cause another (Y), if X is *necessary and sufficient* for Y. It means that X must always be connected to Y; Y must not occur without X, and Y must not need any other phenomenon to occur. But it is sometimes possible to speak of causality when X is necessary but not sufficient for Y. For example, better public service may be associated with an increase in citizen’s trust in the administration, but only in connection with publicity campaigns, favourable media treatment, and sound personnel management in the agencies. An even “weaker” form of causality is found when it is not known whether X is necessary or sufficient for Y, but they tend to be present or absent together. Here are three forms of causality, usually called deterministic causality, probabilistic causality, and correlative causal-

¹³ Danida 2000. Evaluation of Danish Support to promotion of Human Rights and Democratisation 1990–1998: Vol 1. Synthesis Report. Quoted in Crawford and Kearton (2002), page 78.

¹⁴ Sida Stuetes in Evaluation 00/03. The evaluability of Democracy and Human Rights Projects: A logframe-related assessment. ITAD Ltd in association with the Overseas development Institute.

ity. The correlative causality may not involve any causality at all in the common understanding of that word, but as a scientific term it extends the concept of correlation into the explanation of events.

Non-linearity and complex systems

It is commonplace to point to the complexity of development issues, and everybody knows that social change is multidimensional. However, when it comes to defining and measuring results, all this wisdom is forgotten, and surprisingly many are stranded in simple deterministic and mechanistic models of the world. Uphoff (1992) point to four fallacies in analytical orientations when analysing development issues:

1. the reductionist fallacy, which is a dominant strategy in many disciplines, driven by methods of analysis, meaning to simplify phenomena or relationships, casting them into simple either – or categories, or simplistic scales of measurement;
2. the individualist fallacy, treating social or collective phenomena as they were only reflections of personal interest;
3. the materialist fallacy, denying the reality and importance of nonmaterial factors;
4. the mechanistic fallacy, regarding things as if they were machines.

Uphoff calls for a “*post-Newtonian social science that looks beyond reductionist thinking to explore collective action and non-material realities*”. The methods and assumptions of positivist social science do not do justice to values, ideas and motive forces like human solidarity. As these have very real consequences, they deserve more attention than received within reductionist frameworks for modelling the social universe.

But if we leave reductionist and mechanistic models of social systems to the side, what does it actually mean to think of society as a complex, adaptive and non-linear system? Does it make any difference? Yes it does, because in linear models it is, for example, taken for granted that the extent of an effect is similar to the extent of its cause. But in a non-linear system, a tiny event – or a cause – can produce huge, unexpected effects. In popular science this is often referred to as the butterfly effect – a term borrowed from the meteorological sciences. The effects of a butterfly flipping its wings cause tiny changes in air pressure that escalate through feedback processes and may generate a storm on the other side of the earth. It is a popularised image but it is real. There can be no doubt that “butterfly effects” can be found in political developments as well.

Other features of non-linear system are cluster effects and qualitative leaps. The idea of cluster effects means, simply speaking, that all things have to be in place before a system changes. It is like the slot machine where you can only win if the three cherries show at the same time. Similarly, progress in participation, will only happen if several conditions are fulfilled simultaneously, and then progress may suddenly be very fast.

The dominant notion in development theory is that social change occurs gradually and incrementally. Growth is measured by steady increases of a few percentages each year. This cannot be denied, but important changes also occur in leaps and bounds. The situation in a social system may be characterised by some equilibrium, and then there are radical changes until a new equilibrium is reached, and the situation is stable for many years, until a new qualitative leap occurs. Longer views of history seem to bear evidence that systems change in such leaps, rather than through gradual transformations.

So, if we consider that evaluation methodology has to deal with butterfly effects, cluster effects, and qualitative leaps in non-linear systems, does that mean it is impossible to assess results. Definitely not,

in fact social sciences have made a great deal of progress since these concepts were introduced. The development of chaos theory and the sciences of complexity have meant that social science has been renewed, and has been able to address problems in a fashion that is more realistic, more relevant, and hence also more useful. But what does it mean to an evaluation, and how can evaluators work against this background?

Structuring the inquiry with models

One of the challenges for an evaluator is to structure the inquiry. Terms of reference can be of some help, but they are set to define the purpose and to facilitate commissioning the evaluation. Their primary task is not to structure the inquiry in the operational sense of that word. The most important tool that can be used to structure an inquiry is a model of the phenomena to be studied. Models are representations of states, objects, and events. They are idealised in the sense that they are somewhat less complicated than reality and hence easier to use for evaluation purposes, but in a model one can still express contending forces, show cluster effects, and describe qualitative leaps. Evaluations of democratic governance enter a field with complex causal relationships, and it is likely that there will be non-linear dynamics in the interaction. A good descriptive model of the relationships between events is must!

Following the models, another very useful approach to structure an inquiry is to develop a set of hypotheses concerning the developments that may be expected. This helps the evaluator define a focus, to discuss the relevance of that focus, and to assign priorities in the choice of methods. By formulating hypotheses early on in the evaluation, the evaluators will have an obvious instrument to assess whether their thinking is relevant and whether it will generate any new lessons. It leads to an economical use of the evaluation resources. The evaluators can then focus on whether to prove or disprove the hypotheses¹⁵. And if the issue is trivial this soon becomes obvious.

When I now turn to the choice of methods, I will primarily deal with methods of data collection and analysis. In fact, several methodological choices are made even before that, as for example whether or not to develop models or to use hypotheses. There are other than strictly methodological choices, as for example whether to involve stakeholder groups in participatory evaluation). There are those who prefer to start with an overall choice between quantitative and qualitative methods.

The methods of investigation

In essence, an evaluator has a choice between four basic ways of collecting data; interviews, surveys, studying documentation, and observation. The aim of a discussion of methodology would be to ensure that evaluators make flexible, sophisticated and adaptive methodological choices. The following notes introduce the nature of choices, but at this stage it is nothing more of an indication about what a manual of evaluation needs to explain and deliberate on.

1. Surveys and Questionnaires are probably employed more often than any other technique in evaluation. It is not difficult to write questionnaires, but one does need some clear thinking about the kind of information that is needed and an understanding of what kind of questions to ask to get the needed information. Questionnaires probably generate more worthless data than any other technique in evaluation. Greater attention to a few fundamentals of questionnaire writing could improve many evaluations. But in development cooperation, it is rare that evaluators use surveys – presumably because of the difficulties involved in sampling, and practically collecting the data.

¹⁵ This should not be confused with statistical hypothesis testing, which is an altogether different exercise. Statistical hypothesis testing is by many seen as a hallmark of positivist social science, but here I use hypothesis in the manner of much scientific work in the business administration and organisation theory where it is a rhetorical device.

2. An evaluator interviews people to find out from them those things that he or she cannot directly observe. The purpose of interviewing is to allow us to enter into the other person's perspective. Interviewing thus starts from the assumption that the perspective of others is meaningful, knowable, and able to be made explicit. We interview to find out what is in and on someone else's mind, to gather their stories (Patton, 2002). The quality of the information obtained during an interview is largely dependent on the interviewer. Perhaps more than with surveys, interviewing is an art that requires some basic skills – plus plenty of experience.
3. What people say in interviews and express in questionnaires is actually the most common sources of information in evaluations. However, we all know that people do not always do as they say. To understand fully the complexities of many situations, direct participation in and observation of the phenomenon may be the best method.
4. Finally we have the study of documentation; written materials such as program records, memoranda and correspondence; official publications and reports; diaries, letters and artistic works; annual reports, budgets, expenditure accounts, book-keeping. These are subject of different kinds of analysis; they form the basis of quantitative data – you count and calculate efficiency rates on the basis of cost information in expenditure records. But this category of evaluation methods also includes qualitative analysis of texts, for which there is now a variety of analytical tools.

Measurement and scales

Measurement and scales are considered at the same time as the choice of methods is done. I would just like to point at some of the issues that must be considered when the choice of methods is done. The word measurement stands at the centre of attention. It is of course possible to assess progress in democracy, or in the field of human rights, without measuring them. Measurement must be understood as the procedure by which the evaluators obtain symbols, which can be used to represent the concept to be defined (Ackoff, 1962).

The purpose of measurement is to represent the content of observations by symbols, which relate to each other in the same way that the observed objects, events, or properties are or can be. Ranking involves what is known as an ordinal scale. There are those who limit the word measurement to operations that involve an interval or ratio scale. However, it is more common to speak of measurement in a wide sense, involving both ranking and assessments on nominal scales.

The design of the scales raises a number of interesting issues. Is it better to have an odd or an even number of alternatives? Should there be a directly negative option, or is the worst possible outcome that institutional development is “negligible”? If direct value expressions are used, for whom is the outcome deemed satisfactory? What are really the differences between “high” and “substantial”, or between “modest” and “negligible”? Which forms of bias in reporting are inherent in the scales used?

It is worth noting that there is no scientifically agreed ideal of how a scale should be constructed. Sociological research often uses ranking scales where respondents report of something being satisfactory or not. Opinion polls of various kinds do the same, as well as studies exploring corporate cultures. It is also quite possible to assess project efficiency and effectiveness in terms of how satisfactory it is.

But one should bear in mind that the question of whether an activity has been satisfactorily completed or not can be debated, and defined, in relation to whether objectives have been reached, targets met, etc. The credibility of the statement whether something is satisfactory or not rests on the assumption that it is fairly clear why it is satisfactory. There must be some norm that both the evaluator and his/

her audiences can refer to. The statement of whether that impact is satisfactory or not thus risks becoming very personal, and the validity of the observation can be very low.

At this point we can return to the subject of performance rating. The rating systems applied by most agencies in development cooperation¹⁶ build on the judgements of programme officer. The rating reflects his or her opinion on how well a project is doing, whether it will reach objectives, whether it is satisfactory.

Can you trust their assessment? Well that depends on the design of the system. If the people who rate have taken part in the design of the system, if the rating system is accepted as a useful tool by them, if they apply similar standards of what constitutes success and risk, then it can be quite reliable. But if the system is cumbersome, covers very different topics, and – worst of all – is connected to personnel management and career planning, then the data can be highly biased. In most of the organisations that have such system, I would think that the information is not really to be trusted.

From the managerial point of view, it can of course be interesting to know what the staff members think about their projects. But the rate does not say whether the project really is successful or not – for that, you need a far more penetrating analysis than a rating system provides. One has to accept that the rate in a rating system is an expression of opinion – no more, no less. Whether it is worth the cost to solicit such opinions depends on the organisation and the design of the system. When one considers the costs, the most important is probably the opportunity cost of the investment in a rating system.

Quality of evaluation

So the question is, where should the money for evaluation go, and how should it be spent? If you invest in an expensive evaluation process, you have every right to expect a high quality of results. But what is quality, and what distinguishes a good evaluation from a bad one? The question needs some attention. In the past, quality was presumably assured by adhering to scientific discipline. But with the publication of the Programme Evaluation Standards (1994) things changed. There is now a commonly agreed standard of what quality is, and what properties a good evaluation process should have. The Standards were established by the American National Standards Institute as approved Standards. The actual content was developed over several years by the Joint Committee on Standards, which was established by the American Evaluation Association, and a wide number of professional associations, consultants, academicians, and public authorities.

The Standards are used in professional contexts as a guide to what constitutes a reasonable evaluation. The Standards are intended to reflect the best practice in the international evaluation community, and were set in order to encourage useful, feasible, ethical and sound evaluations, which in turn will contribute to the betterment of social interventions in diverse settings. The Standards were not established for any particular type of evaluation, so they are as relevant for evaluation of citizens' participation as for anything else.

The Standards are organised around four important attributes of an evaluation: *utility, feasibility, propriety, and accuracy*. These four attributes are necessary and sufficient for sound and fair evaluation. We should recognise that the standards relate to each other. An evaluation that is not feasible is not likely to yield accurate conclusions, and conclusions that are not accurate are not likely to be used. Similarly, an evaluation that is conducted according to high standards of propriety will generally have much higher utility than one with shortcomings in these respects (Burke Johnson, 1995).

¹⁶ Comprehensive ratings systems are used by the multilateral development banks, the EU, and KfW. Sida has developed a ratings system, and it is being put to wider use at present.

Utility standards guide evaluations so that they will be informative, timely and influential. They require evaluators to acquaint themselves with their audiences, define the audiences clearly, ascertain their audiences' information needs, plan evaluations to respond to these needs, and report the relevant information clearly and in a timely fashion." I chose to put utility at the top, and so did the authors of the standards. This is no coincidence. Evaluations have a mandate to be useful, and if not put to use – in one way or another – they were a waste of time and effort. This is the distinguishing mark, and what makes evaluation different from social science research in general. It is the most important quality!

Accuracy standards determine whether an evaluation has produced sound and trustworthy data, leading to valid and reliable conclusions. The evaluation of a project must be comprehensive; that is, the evaluators should have considered as many of the programs features as practical and should have gathered data on those particular features judged important for assessing the project's worth or merit. Moreover, the information must be technically adequate, and the judgments rendered must be linked logically to the data.

Feasibility standards recognise that evaluations usually are conducted in a natural, as opposed to a laboratory setting and consume valuable resources. Therefore evaluation designs must be operable in field settings and evaluations must not consume more resources, materials, personnel, or time than necessary to address the evaluation questions.

Propriety standards reflect the fact that evaluations affect people. These standards are intended to protect the rights of individuals. They promote sensitivity to and warn against unlawful, unscrupulous, unethical, and inept actions by those who conduct evaluations. The standards require that individuals conducting evaluations learn about and obey laws concerning such matters as privacy, freedom of information, and the protection of human subjects. They charge those who conduct evaluations to respect the rights of others.

The standards apply to all kinds of evaluation, but the risks are not equally large – or of the same kind – in all fields. What are then the particular threats to quality in development cooperation – and particularly in respect of evaluation of democratic governance? The ethical issues are very complex. An evaluator has to pay far more attention to issues of representation than is normally the case, and in the process of evaluation one has to be aware of the hidden stakes and interests of all partners. Depending on context, the issue of anonymity can be very sensitive. Accuracy is always an issue, but it is particularly difficult when the evaluation has to assess facts, opinions, values and attitudes. The one fuses into the other, though they are very different empirical realities. In the course of an interview situation, or in a survey, the respondent may become as confused as the evaluator.

The quality standards are important steering instruments. If evaluators are aware of the criteria their products will be assessed against, it is likely that they will be more concerned about the quality of their products. Disseminating quality standards and using them in follow-up processes are good ways of making sure that the money allocated to evaluation is well spent.

The discussion suggests that it is both necessary to pursue the quantitative measures of progress that are provided by organisations such as Afrobarometer, World Value Surveys, Freedom House, Inter-Parliamentary Union, Transparency International, etc. Independent bodies should gather such statistics. Neither that nor theory should be part of any system of performance management. With the aid of data provided by independent organisations, it is possible for the aid agencies to commission evaluations in direct connection to programme delivery, and to expect the evaluators to come up with high quality studies; studies that build on indicators. These indicators are some of many sources of data, and they must be put in a proper framework of qualitative and quantitative analysis.

Concluding remarks

In spite of the often-quoted difficulties in establishing causal connections between programs in development cooperation and social changes in democracy and governance, it is quite possible to produce relevant and interesting conclusions in evaluation reports. However, the evaluators must avoid the four fallacies mentioned here (the reductionist, individualist, mechanistic and materialist fallacies) and to do so they need to take account of social systems as complex and non-linear entities. In practical terms, that means evaluators have to pay more attention to context, and to structure enquiries with the help of models. Delving into complex systems can otherwise be a very confusing experience, if not for the evaluator, then certainly for the reader of a report.

In terms of performance management, it is highly unlikely that any one evaluation will present results in terms that are comparable to any other evaluation. Hence it is impossible to build any rating system on measurements that come out of the reality of the programs. Most rating systems build on subjective opinions on performance that may be interesting in their own right. However, the usefulness lies primarily in day-to-day management, and it is not to be confused with an overall reporting of results. It is thus highly unlikely that one can report on progress against the development objectives in general terms. One can report specifically, within the context of one evaluation and one program, but no more.

If evaluations are to be used to account for results, it is of course essential that they are good. A good evaluation is characterised by utility, propriety, validity and feasibility. Unless evaluations possess these attributes, they will not account for results, and hence will they not contribute to organisational accountability.

Chapter 5. How do we evaluate democracy “at home”?

Can democracy be evaluated?

Is it really possible to evaluate such a complex and nebulous concept as democracy? Of course it is, and it is done all the time! At the time when this report is written, there is less than a month left to the election day in Sweden, when we cast our votes for Parliament, as well as for regional and municipal representative bodies. Hardly a day passes without someone proclaiming the decline of political parties, the demise of public debate, the catastrophically low level of public interest in political affairs etc. Almost equally often someone will point to new forms of public engagement, the virtues of present forms of civic engagement, the development of local democracy – all indicating that the present political system works rather well.

In 1985, the Government appointed a public inquiry into the distribution of power and democracy in Sweden¹⁷. The terms of reference express a worry that social development call for a renewal of political systems, and that there is a need to systematically assess the challenges to the democratic ideals, and to develop initiatives to make the access to power more equitable. The committee started working in 1985 and handed over its main report to government 5 years later, in June 1990. The inquiry was led by a team of four scientists, but also relied on large number of associated researchers. The final report answers the questions above on 411 pages¹⁸. In the course of those 5 years, another 28 books were pub-

¹⁷ Utredning om maktfördelning och demokrati i Sverige. Dir 1985:36.

¹⁸ SOU 1990:44. Demokrati och Makt i Sverige. Maktutredningens huvudrapport.

lished, plus 34 papers, technical reports and drafts (in Swedish), another 43 papers etc. in English, plus 13 progress reports, conference proceedings, and the like. All in all, it took (my estimate) some 30 to 40 working years, resulting in some 15.000 pages of text, to respond to the questions asked by the government – to assess the state of democracy and the distribution of power in Sweden. And there is no doubt that this is a thorough assessment.

But the reports raised as many new questions as it answered old ones. It did not last long until the Government appointed a new public inquiry into the state of Swedish democracy. “Demokratiutredningen” started working in 1997, and presented its final report in the year 2000¹⁹. The purpose was to analyse and discuss the conditions for democracy in Sweden in the next century, and to identify problems and opportunities for the public participation in political processes²⁰. The new inquiry was not only expected to throw light on these issues, but it had a mandate to engage society in the discussion, and to construct a platform for an intensified debate on the development of democracy. The work was not only expected to produce research reports, but it was to spearhead a wide-ranging debate in society. The final report is only the tip of an iceberg. There were many other activities that do not lend themselves to easy integration in a final report.

The two public inquiries shared the interdisciplinary approach, and they both devote considerable time to the theoretical aspects of power, democracy and participation. The main factor, which distinguishes the two processes, was the participatory nature of the latter. The advances in technology had made a totally new form of involvement possible, and this was used in the design of public meeting place (“demoikratitorget”). This was web-based forum for discussion, which fed into the inquiry through formulating questions, opinions, and providing empirical evidence on the subject.

These two public inquiries show that it is not only in developing countries that there is some concern for how and why society makes progress in a democratic direction. Similar question are posed and answered in the domestic debate, but as these examples show, it takes a good deal of work to provide adequate answers. Theoretical as well as practical issues are considered, and it takes time to get answers. The process of inquiry must be interdisciplinary, participatory, and must be allowed to take time. As knowledge grows, there are new questions to be answered in a never-ending pursuit of good governance.

Evaluating democratic institutions

Swedish legislation, as well as many other forms of political initiative, starts in the system of public inquiries. Parliamentary inquiries are one of the fundamental institutions in society; it is expected to generate high quality proposals for legislative development, to generate penetrating insights into any subject that Parliament and Government ask to have investigated, and the system as such is to be transparent and open, as well as to work effectively, efficiently, and in quick response to the needs of government.

There is a debate on how well the system of public inquiries actually works. The Parliament’s Auditors published a critical examination in 1996, which pointed to problems in respect of the evaluation competence, time frames and budgets for committees, fuzzy mandates as well as limited consultations around completed reports, and poor dialogue with the government ministries²¹. Two years later, a com-

¹⁹ En uthållig demokrati! Politik för folkstyrelse på 2000-talet. SOU 2000:1

²⁰ Direktiv 1997:101

²¹ Riksdagens Revisorer, Rapport 1996/97: 6

mittee under the Ministry of Finance undertook a new study of the quality of the system of public inquiries²². This was a thoroughly conducted evaluation, based on a stratified random sample of 20 reports (each year, around 100 committees public the results of their work, and at any point in time, there is usually some 300 committees at work).

The evaluations have been carried forward²³, and at present there is a system of quality control – an ongoing, internal evaluation system – being designed. The evaluation system builds on a model, which specifies a number of quality attributes. The model is operationalised through a standard format of 105 variables, and when it is applied, each committee report is “graded” in respect of each variable. The numbers from the checklists are entered into a database where it is possible to subject the data to a number of analytical tools.

Evaluating governance – civil service reform

Good governance is a goal in all societies, and it is one of the major areas of intervention in development cooperation. Projects and programmes in this field have as their objective to increase responsibility, accountability, transparency, participation, predictability and efficiency of organisations in the public sector²⁴. It would go much beyond the scope of this paper to present all projects in Sweden that aim to improve the performance of the public sector. There is hardly a public agency, which is not undergoing some form of organisational change, and in the large public administrations, there are usually many such initiatives running in parallel at any one point in time. The same holds true for most other countries; there is hardly a government agency, which is not touched by sweeping changes of reform, and many of these are also subject to evaluation. Indeed, the last annual meeting²⁵ of the European Evaluation Society devoted the whole program to the politics of evaluation – and the evaluation of politics.

New forms of representation and public participation are emerging in all countries. These developments have expanded the avenues for citizens to participate more fully in public policy-making, within the overall framework of representative democracy. Citizens are increasingly demanding greater transparency and accountability from their governments, and want greater public participation in shaping policies that affect their lives. In 1999, the PUMA Working Group on Strengthening Government – Citizen connections launched a survey of the legal, policy and institutional frameworks in OECD Member countries to ensure citizens’ access to information, consultation and active participation in policy-making. A second survey followed in 2000 on the use of new information and communication technologies in strengthening government – citizen connections. Following that, a set of country case studies highlighting concrete experiences in some member countries was produced. These surveys led to the publication of the OECD report “Citizens as Partners: Information, Consultation and Public Participation” (2001) and accompanying handbook²⁶.

²² ESO (Expertgruppen för Offentlig Ekonomi, the Ministry of Finance): “Kommittéerna och bofinken – Kan en kommitté se ut hur som helst?” (Ds 1998: 57)

²³ see for example; Konstitutionsutskottet; KU 31 1997/98; Sten Johansson, Kvalitet och effektivitet i kommittéernas arbete, Förvaltningsavdelningen (RK/FA 1998.10.20), and the Governments Policy for the Reform of Public Administration (Förvaltningspolitiska Handlingsprogrammet).

²⁴ Sida’s activities in the sector are described in the internal working paper “Good Governance” from the Division for Democratic Governance.

²⁵ The conference was held in Lausanne in October 2000, and the proceedings, keynote speeches, as well as all papers (in pdf format) can be seen at the website of the European Evaluation Society (www.ees.com)

²⁶ The books are available in paper cover, but they can also be downloaded from the OECD website; www.oecd.org

The report devotes one section to “developing the capacity for evaluation”, but as I mentioned, the report contains a number of case studies. Each of these can be seen as a “mini-evaluation”. The cases describe the projects, explain purpose and context, tell the story of how the projects were implemented, and analyses the results. What more could one ask for? Well, the PUMA Working Group noted that so far, little systematic evaluation had been conducted on the different activities meant to strengthen government – citizen connections, particularly then with the use of information and communication technologies. In response to that, it was decided to develop a framework of evaluation, so as to guide the member countries in their efforts to evaluate projects in this field.

Evaluating human rights

At first we may not think that there is human rights evaluation in Sweden. Who would do that and why? Those kinds of evaluations would certainly be relevant in Africa, on the Balkans, perhaps even in the Baltic states in connection with the status of immigrants. But in Sweden? Well, there are several examples of situations where it is necessary and urgent to evaluate how human rights are treated in Sweden too. As one of the many countries that have ratified the Convention on the Rights of the Child, the Government of Sweden is expected to present regular reports to an international commission on how it works to make sure that children’s rights are indeed protected. And when the Government presents its report, the Swedish section of “Save the Children” (Rädda Barnen) presents an alternative report.

The report from Rädda Barnen is a comprehensive evaluation, building on a joint effort from local sections, penetrating specific issues of how vulnerable children are treated in Sweden. It is a detailed report, building on several months, or indeed years of work, mainly by volunteers and idealists. The main methodological choice is the interview. There are many excellent examples of good interviews, both with individual respondents and in groups. There is less of quantitative approaches, and not much use of statistics or other aggregate data. The report²⁷ does not venture into any overriding conclusions about the welfare of children in Sweden, but it points to a number of very practical and concrete examples, where Swedish society, while still being relatively affluent, could be expected to comply better with the universal obligations.

Concluding remarks

So what lessons could be learnt from “domestic” evaluation? First, it is indeed possible to evaluate progress in all these respects, and it is done all the time, but not in terms of accountability for results. Second, evaluations take time and need resources, and they require interdisciplinary teams and methodological pluralism. Third, most evaluations take a participatory character. Fourth, evaluations often have a practical focus on how to improve practical aspects of governance, democracy and human rights. Finally, evaluations enlighten the debate, guide decision-making, extend knowledge, but they do not provide definite answers in respect of complex and multidimensional issues such as the state of governance in society.

²⁷ At the time of writing, “Rädda Barnens alternativrapport” is not yet made public, but should be available under the website of Rädda Barnen; www.rb.se in the near future.

Chapter 6. Conclusions and recommendations

Common frameworks?

Are there any common models or common approaches? Crawford and Keaton (2002) did a major survey of the practices of bilateral and multilateral donors. They did indeed find that there were many evaluations, far more than they had expected to find. The database they built consists of 110 references. I would think that there are far more evaluations, and Crawford and Keaton also acknowledge that there might be more. As the broad field of democracy and human rights include governance, which again includes all forms of support – technical assistance, twinning arrangements, sector programs – to public administration, there are indeed many more evaluations available.

The answer to the above question depends a little bit on what we mean with models. The problem is that, for instance, Webster's New Collegiate Dictionary lists twelve different meanings of the word "model". Sifting out the most obviously inappropriate ones, there are still quite a few possibilities left²⁸:

"...a miniature representation of something; also a pattern of something to be made; an example for imitation or emulation ..., archetype, a description or an analogy used to help visualize something that cannot be directly observed."

I have used the word "model" in previous chapters in the latter sense (a description or an analogy ...). Taking that as a starting point, there are few common models in the field of evaluating democracy and human rights. But what does that mean, and what is it that we are looking for models of? There are at least three possible answers to that question.

Models for accountability

First, there could be models in respect of how organisations set up their systems for accountability. Such systems can, as we have seen, be composed of evaluation systems or performance management systems. An evaluation system has many design variables²⁹, and there are numerous ways to balance integrity, funding, personnel policies, etc. Accountability systems could build on combinations of performance management and evaluation.

If the word model is used in this sense, there are certainly not any models that are commonly used among funding agencies. Each has its own system of reporting, which is only to be expected. Each bilateral funding agency reports to a minister, a Government and a Parliament, steeped in a particular national administrative culture of control and decision-making. They face different audiences, and of course they have to comply with national traditions.

It is commendable that the OECD/DAC working group on aid evaluation has managed to achieve some common approaches to definitions, purpose and policies in evaluation. It would have been even better if the definitions of the group had been aligned with the main currents in evaluation research, as that would facilitate exchanges of experience etc. between evaluation on the domestic scene and evaluation in development cooperation. Even if there are no common models yet, there are concerns that are common to all, and that may lead to joint/coordinated/coherent approaches in the future – if not common models.

²⁸ The example is fetched from Barbara Czarniawska: *Writing Management: Organisation Theory as a Literary Genre*. Oxford University Press, Oxford. 1999.

²⁹ See for example Forss, 2001, which is a report to the SOU 2002:1.

Models for evaluation systems

Second, the word model could also be used to describe how the agencies go about commissioning evaluations; that is, how do they practically get evaluations done. Donor agencies rely on external consultants and researchers. Their evaluation process goes from initiating the work, formulating terms of reference, recruiting evaluation teams, monitoring the teams work, and receiving reports and putting these to use.

There are many similarities between agencies, so similar are they that one could actually speak of a common model. The way they work in this field is eloquently described in different evaluation manuals. But remember that these manuals are for evaluation management (from the agencies' side) rather than manuals on how you do the dirty work of the evaluation.

Models in evaluation

Third, in respect of actually doing the evaluation, each team (or evaluator) makes its methodological choices, designs its own instruments, analyses according to its preference for structure and logic. Few, if any, use models in the practical execution of the task, although there would be much to be gained by that. Even if a particular evaluation team is willing to learn from another, and seeks inspiration in previous reports that have been well received, the practice is to develop new approaches. The fact is, that an evaluator, who simply used the models of another evaluator, would be accused of plagiarism and would get a bad reputation.

There will never be any common models to guide the practical evaluation work (but evaluations teams could still benefit from using models). There is no evaluation task that is quite like any other, hence you cannot work the same way twice. In addition, even if there are good evaluations, I have neither seen one, nor done one, which could not be improved. It lies in the nature of evaluation that you experiment and work with limited resources. I would think there are few evaluators who, when looking back at a process of evaluation, do not see things that could and should be done differently next time a similar job comes along.

Approaches to accounting for results

There are funding agencies that account for results in respect of democratic governance by performance management systems, notably USAID, KfW and the multilateral development banks. Their systems have been criticised by the research community, and it is an open question whether their systems say much about actual results.

Most agencies rely on traditional evaluation systems, but that is also an approach that has its problems. First among these are;

- The difficulty of presenting aggregate results based on different, context specific evaluation reports
- The fact that it takes time and resources to account for the complexities in progress towards democratic governance
- Those on the receiving end of the information, do not have the patience and the understanding to read and digest research cum evaluation texts
- The quality of evaluation reports can be a problem, as they do not always present reliable and valid findings

In order to improve the account for results, there are thus two choices; either to solve the problems inherent in the present evaluation system, or turn to new forms of accountability, that is, the kind of performance management systems based on ratings and indicators that were discussed above.

The conclusion here is that there is more of a need for comprehensive, contextually rich, and thorough understanding of democratic governance, than there is for simplistic and standardised approaches. It would thus be more cost effective, and provide better results in the end, to develop evaluation systems to account for results in democratic governance.

Approaches to indicators

Indicators are important in most research as well as in evaluation and in systems to account for results. Indicators have a role to play at several levels in the analysis of results.

- To describe the overall trends in democratic governance. So for example the indicators developed by the UNDP, Freedom House, Transparency International, the World Bank, etc. can all be used as inputs to a discussion of developments.
- Indicators can also be part of the analysis in evaluations, and most evaluations do in fact work with a number of indicators, based on information from interviews, surveys, documents etc. But the systems of indicators are unique to each evaluation, and must be so in order to be accurate and cost-effective.
- Furthermore, it is possible to use indicators to describe results in respect of plains and simple objectives, as for example, the delivery of services, expenditures, and basic activities. Even if these kind of results are less interesting, and far removed from the goals and purposes of democratic governance, they have a role to play in a comprehensive analysis of results

Approaches to objectives

Clear thinking and practical action, based as always on inadequate evidence, are all we have to work with in the field of democratic governance as well as in other areas. It is obviously good to be clear about what one tries to achieve and why. However, it is equally well known that it is practically useful to be flexible, and to be able to adjust objectives along the way. Hence, a blueprint approach to planning, the design of logical frameworks, and the hierarchical ordering of objectives is not necessarily synonymous to clear thinking.

This study does not treat the design of objectives, but in general, it would seem as if the final results of development cooperation are more important than the formulation of plans and objectives. It is not practically useful to spend huge amounts of money and resources on framework exercises that have limited relevance and that are regularly overtaken by events. There is an opportunity cost to the investment in planning systems, which is considerable.

It is sometimes claimed that clear objectives is a prerequisite for evaluation and for any type of account for results. I would like to end this paper on the optimistic note that evaluation is always possible and can always be relevant. There are even many who claim that goal-free evaluation is the most appropriate approach of all, hence suggesting that any knowledge of objectives introduces a bias in evaluation. That may be taking the point a bit far, but more importantly, systematic evaluation needs to go beyond objectives, goals and purposes in order to arrive at a comprehensive analysis of worth and merit.

The road ahead

It is not the task of this paper to present detailed recommendations on how to develop the system to account for results in respect of democratic governance. However, there are some common themes that emerge out of the analysis, and that point out a general direction of the efforts. So, what are the main suggestions to be derived from this study concerning how to account for results in the field of democratic governance:

1. Put the emphasis on evaluation research, which is the instrument that allows the organisation to handle complex and multidimensional phenomena.
2. Develop the quality of evaluations by emphasising the need for modelling the research context, structuring the inquiry with the help of relevant hypothesis, and encourage methodological pluralism.
3. When allocating resources, consider that it is better to do some few evaluations of high quality than many with questionable validity and reliability.
4. Learn from domestic experience that proper evaluation research takes time and money, if it is to be useful.
5. Develop the communication with the audiences who are concerned with results, and address their needs for short, relevant, and easily digestible information based on evaluation findings.
6. Performance management systems can be applicable when accounting for results at lower levels, mainly relating to expenditure items and descriptions of activities, but they are not appropriate for analysis of results at the level of goals and purposes.
7. Indicators are tools that can be used at several levels: (1) to outline development trends, (2) to provide inputs to the qualitative analysis, and (3) to describe results at the level of activities. There is a need to develop the methodological competence to work with indicators at each level, according to the role that can be expected of the instrument at that level.

References

- Ackoff, R. (1962) *Scientific Method*. New York: John Wiley:
- Berg, E., 1994, *Rethinking Technical Cooperation: Reforms for Capacity Building in Africa*. New York: UNDP and Development Alternatives Inc.
- Berlage, Lodewijk and Olav Stokke, eds., (1992) *Evaluating Development Assistance: Approaches and Methods*, London: Frank Cass.
- Blalock, A.B. (1999) Evaluation Research and the Performance Management Movement: From Estrangement to Useful Integration? *Evaluation*, Vol 5. No 2. Sage, London.
- Boyle, R. And Lemaire, D. (1997) *Building Effective Evaluation Capacity: Lessons from Practice*. Transaction Publishers: New Brunswick.
- Burke Johnson, R. (1995) Estimating an Evaluation Utilisation Model Using Conjoint Measurement and Analysis. *Evaluation Review*: 19, 3, p 313 – 338.
- Carothers, T. (1999) *Aiding Democracy Abroad: The Learning Curve*. Carnegie Endowment for International Peace, Washington, D.C.
- Cassen, R. and Associates, 1986, *Does Aid Work?* Suffolk: Clarendon Press.
- Cracknell, B. (2000) *Evaluating Development Aid: Issues, Problems and Solutions*. Sage: London..
- Crawford, G. and Kearton, I. (2002) *Evaluating Democracy and Governance Assistance*. Centre for Development Studies, University of Leeds
- Dahler-Larsen, P. (1998) Evaluering, legitimitet og inkonsekvens. In Schaumburg-Muller, H. And Wad, P., ed. (1998) *Evaluering af udvikling*. Den Nye Verlden: Copenhagen.
- Davis, I.C. (1999) Evaluation and Performance Management in Government. *Evaluation*, Vol 5. No 2. Sage, London.
- De Vaus, D. (2001) *Research Design in Social Research*. Sage: London.
- DsUD 1990:63. *Bra Beslut: Om Effektivitet och Utvärdering i Biståndet*. Utrikesdepartementet: Stockholm.
- Forss, K. (1985) *Planning and evaluation in aid organisations*. EFI/IIB: Stockholm.
- Forss, K. and Carlsson, J. (1997) The quest for quality – or can evaluation findings be trusted? *Evaluation*, Vol 3, No 4.
- Forss, K., Carlsson, J. And Rebien, C. (2002) Process Use of Evaluations: Types of Use that Precede Lessons Learned and Feedback. *Evaluation*. Vol 8. No 1.
- Forss, K. (2001) The theory and practice of analysing institutional development in evaluation. *A report commissioned by the ECG* (Evaluation Consultation Group of the Multilateral Development Banks)
- Joint Committee on Standards (1994) *The Program Evaluation Standards*. Sage: London.
- Kvale, S. (1996) *Interviews: An Introduction to Qualitative Research Interviewing*. Sage: London.
- Lonner, W.J. and Berry, J.W. (1986) *Field Methods in Cross Cultural Research*. Sage: London.

- Lusthaus, C. et al (1999) Enhancing organisational Performance: A toolbox for self assessment. IDRC, Ottawa
- Majone, G. (1989) *Evidence, Argument and Persuasion in the Policy Process*. Yale University Press: New Haven.
- Maren, M. (1997) *The Road to Hell: The Ravaging Effects of Foreign Aid and International Charity*. The Free Press: New York.
- Mintzberg, H. (1994) *The Rise and Fall of Strategic Planning*. New York: The Free Press
- Mosley, P. (1987) *Overseas Aid: Its Defence and Reform*. Weathersheaf: Brighton.
- OECD, 1986, Methods and Procedures in Aid Evaluation. OECD, Paris
- OECD, 1991, Principles for New Orientations in Technical Cooperation. Paris.
- OECD, 1995, Participatory Development and Good Governance. Development Cooperation Guidelines Series. Paris.
- Patton, M.Q. (1997) *Utilization-focused Evaluation, 3rd Ed.* Sage: London
- Patton, M.Q. (1982) *Qualitative Research and Evaluation Methods*. Sage: London
- Pawson, R. and Tilley, N. (1997) *Realistic Evaluation*. Sage: London
- Rebien, C. (1996) *Evaluating Development Assistance in Theory and Practice*. Avebury: Aldershot.
- Riddell, R.C. (1987) *Foreign Aid Reconsidered*. The Johns Hopkins University Press: Baltimore.
- Samset, K., Forss, K., and Hauglin, O., 1993, International Learning from Evaluations and Reviews. *Evaluation Report 1.93*, Oslo: Ministry of Foreign Affairs.
- Scriven, M. (1991) *Evaluation Thesaurus*. Sage; London.
- Stokke, O., ed (1991) *Evaluating Development Assistance: Policies and Performance*. Frank Cass: London
- Thoening, J.C (2000) Evaluation as Usable Knowledge for Public Management Reforms. *Evaluation* Vol 6, N0 2. pp 217– 231
- Thomas, A. Et al (1998) *Finding Out Fast: Investigative Skills for Policy and Development*. Sage: London.
- Torres, R.L.; Preskill, H.S. and Piontek, M.E. (1996) *Evaluation Strategies for Communication and Reporting*. Sage: London
- Uphoff, N. (1992) Learning from Gal Oya. Possibilities for Participatory Development and Post-Newtonian Social Science. Cornell University Press. Ithaca.
- Weick, K.E. and Westley, F. (1997) Organisational Learning : Affirming an Oxymoron. In Clegg et al: *Handbook in Organisation Studies*. Sage: London.
- World Bank, 1996d, *Performance Indicators for Technical Assistance Operations*. A First Edition Note. Operations Policy Department. Washington, D.C.: World Bank.
- World Bank (1998) *Assessing Aid: What Works, What Doesn't, and Why*. Oxford University Press: Oxford

Annex 1. Terms of reference

UPPDRAGSBESKRIVNING

FÖR STUDIE AVSEENDE UTVECKLING AV RESULTATREDOVISNING MOT DET BISTÅNDSPOLITISKA MÅLET DEMOKRATISK SAMHÄLLSUTVECKLING

1. Beskrivning av Sida

Styrelsen för internationellt utvecklingssamarbete, Sida, svarar för det bilaterala svenska utvecklings- och katastrofbiståndet. Sida handhar även huvuddelen av samarbetet med länder i Afrika, Asien, Latinamerika samt Central- och Östeuropa. Sida har omkring 650 medarbetare - inklusive expertis (varav ca 100 i fält) inom ekonomi, teknik, jordbruk, hälsovård, utbildning och miljövard. De övergripande målen för svenskt bistånd är att bidra till minskad fattigdom, ökad demokratisering och uthållig utveckling i de länder Sida har valt att verka.

Sidas avdelning för Demokrati och Social Utveckling (DESO) innefattar huvudsakligen uppgifter som faller inom ramen för de verksamhetsgrenar i regleringsbrevet till Sida som omfattar Demokratisk samhällsstyrning, mänskliga rättigheter och konfliktförebyggande verksamhet samt sociala sektorer. DESOs verksamhetsidé är att genom långsiktigt utvecklingssamarbete skapa förutsättningar för en demokratisk samhällsutveckling och bärkraftig social och kulturell utveckling som gynnar det stora folkflertalet. Avdelningen har som sin huvuduppgift att bereda stöd till insatser inom kultur/media, demokratisk samhällsstyrning, hälsa och undervisning samt utveckla policies och metoder för hela Sida.

Enheten för Demokratisk Samhällsstyrning (DESA) är en av fyra enheter inom DESO, och är Sidas ämnesenhet för Demokratisk samhällsstyrning, med fokus på beredning av insatser och rådgivning inom ämnesområdet. Begreppet "Demokratisk samhällsstyrning" används inom OECD/DACs nätverk för god samhällsstyrning och deltagande. Begreppet har dock inte någon entydig och klar definition men avser täcka områdena; Demokratisering, Mänskliga Rättigheter, God samhällsstyrning, och Folkligt deltagande.

2. Bakgrund

Regeringens regleringsbrev till Sida avseende budgetåret 2001 anger att Sida ska inleda, och i regleringsbrevet avseende år 2002 att Sida ska fullfölja: ett "...metodarbete som syftar till att utveckla resultatredovisning mot det biståndspolitiska målet demokratisk samhällsutveckling." Vidare sägs att "Metodarbetet skall fördjupa problemanalysen, göra tydligare målbeskrivningar och välja indikatorer som ökar mätbarheten över tid." DESA har uppdraget inom Sida att genomföra detta.

Huvudordet i uppdraget är i DESAs ögon "resultatredovisning". Metodarbetet för detta syfte delas upp i tre delar; "fördjupa problemanalysen", "tydligare målskrivningar", och "välja indikatorer", vilket i sin tur avser att "öka mätbarheten över tid".

Uppdraget fokuserar biståndsmålet demokratisk samhällsutveckling, vilket är ett vidare område än DESAs mandat; demokratisk samhällsstyrning.

DESA har under år 2001 och 2002 specifikt arbetat med denna uppgift i regleringsbrevet i huvudsak genom ett antal interna seminarier. Dessa har haft som mål att fördjupa kunskapen inom DESA om resultatredovisning utifrån olika exempel på arbetssätt som är relevant för stöd till demokratisk samhällsutveckling. Initiativet att göra den nu aktuella studien har definierats inom detta arbete.

DESA har under de senaste åren arbetat med metodutveckling inom fyra av DESAs mångfacetterade verksamhetsområde; rättssektorn, politiska institutioner, deltagande och god samhällsstyrning. Ett syntesdokument har utarbetats under 2002 som bl.a. ska utgöra ett underlag för dialog om fortsatt målstyrning och resultatredovisning av verksamheten mellan UD och Sida.

3. Syfte och mål

Studien ska ge underlag för en välgrundad och inom DESA gemensam förståelse för möjligheter och problem med resultatredovisning inom verksamhetsområdet demokratisk samhällsstyrning. Denna gemensamma förståelse ska ligga till grund för hur DESA i givna situationer på bästa sätt kan nyttja och operationalisera olika metoder och angreppssätt avseende resultatredovisning av stödd verksamhet, i syfte att öka lärandet och förbättra verksamheten.

en tydligare och mer informerad dialog med svenska och internationella samarbetspartners och med andra verksamhetsområden inom Sida och på Ambassader avseende resultatredovisning inom området demokratisk samhällsutveckling.

DESAs rapportering i Sidas årsredovisning till regeringen.

För att fylla ovan angivna syfte ska en kartläggning göras där det framgår vad andra biståndsorgan gör, var forskningsfronten ligger, och vad som diskuteras i utvecklingen av ämnet. Uppdraget ska genomföras i form av en litteraturstudie som fokuserar relevanta metoder och begrepp, ger en översikt av kunskapsläget ("state of the art") avseende resultatredovisning inom sektorer relevanta för DESAs verksamhetsområde, Uppdraget ska resultera i en läsbar och pedagogiskt upplagd rapport.

4. Avgränsning

Studien ska fokusera verksamhetsområdet demokratisk samhällsstyrning, vilket i denna studie innefattar den verksamhet som DESO/DESA ansvarar för och som definieras av OECD/DAC som Demokratisering, Mänskliga Rättigheter, God samhällsstyrning, och Folkligt deltagande.

DESAs tidigare nämnda områden för metodutveckling (rättssektorn, politiska institutioner, deltagande och god samhällsstyrning), samt metodutvecklingens syntesrapport, anger DESAs huvudsakliga verksamhetsområde.

Det förutsätts att resultatet av studien, vars avgränsning är demokratisk samhällsstyrning, kan generaliseras till det vidare målet i regleringsbrevet vars fokus är demokratisk samhällsutveckling,

Studiens huvudsakliga informationskälla ska vara dels relevant litteratur, rapporter samt beskrivningar av metoder som explicit berör resultatredovisning för biståndssamarbetet inom området demokratisk samhällsstyrning, och dels relevant akademisk litteratur avseende resultatredovisning. Genomgången av "state of the art" ska endast ta med sådant som kan bedömas som realistiskt och rimligt att kunna operationalisera.

5. Uppdrag

5.1. Studiens tre huvudpunkter är:

Kartläggning av utarbetade modeller.

Att göra en sammanfattande kartläggning och beskrivning av relevanta metoder för resultatredovisning inom verksamhetsområdet demokratisk samhällsstyrning. Kartläggningen ska ge en beskrivning av

respektive metod, dess principiella grunder, samt erkända/möjliga styrkor och svagheter. Kartläggningen ska belysa begränsningar och möjligheter med respektive metod avseende på en fördjupad problemanalys, tydliga målbeskrivningar och val av indikatorer. En annoterad bibliografi (lista) ska göras utifrån genomgången.

Aktuell diskurs

Att göra en kritiskt analyserande sammanfattning av den diskurs som pågår inom området resultatredovisning avseende verksamhetsområdet demokratisk samhällsstyrning. Problemställningar där konsensus råder respektive skillnader finns mellan olika synsätt och aktörer, bör tydliggöras. Sammanfattningen av diskursen ska innehålla en kort tillbakablick men ska framför allt vara framåtsyftande och söka identifiera trender på området. Rapporten bör även inkludera (explicit uttryckta) slutsatser och reflektioner från uppdragstagaren.

Beskrivning av begrepp.

Grundläggande definitioner och beskrivningar av relevanta begrepp inom området resultatredovisning ska återges, samt en beskrivning av hur begreppen används. Beskrivning ska ge verktyg att tolka och bearbeta studiens kartläggningen av modeller och aktuell diskurs.

5.2. Omfattning av studien:

Nedan angivna områden ska ses som värdefulla exempel på områden och aktörer som så långt möjligt ska täckas av studien. Uppdragstagaren ska dock därutöver inkludera så långt möjligt andra relevanta metoder, aktuell diskurs och begrepp. Uppdraget ska i lämpligt sammanhang diskutera regleringsbrevets fokus på "tydliga målskrivningar", "välja indikatorer", samt "öka mätbarheten över tid".

Avseende av biståndsaktörer utarbetade eller påbörjade metoder för resultatredovisning kan nämnas OECD/DAC Governance-projekt, UNDPs "Human Development Report 2000", World Bank andra generationens indikatorer, Raul Wallenberg-institutet, NORAD: Handbook on Human Rights Assessment, Sida/SEKA "Oktagonen", Sidas projekt för "Rejting" inklusive ordinarie projektrapportering till Sida, SMR "Utvärdering – gissel eller möjlighet?", Diakonia "DemEsti", CIVICUS Index, etc.

Studien "Evaluating Democracy and Governance Assistance" (Crawford, Feb 2002, University of Leeds), överlappar delar av det nu aktuella uppdraget och bör tas tillvara.

Sidas Utvärderingsenhets erfarenheter och problemställningar på området (bl.a. Sida Evaluation nr 00/3 samt DESA management response 01-03-08), liksom även den svenska debatten om resultatredovisning inom delar av svensk offentlig förvaltning som har bäring på demokratisk samhällsstyrning, ska tas tillvara. OECDs metodutvecklingsarbete inklusive demokratiutredningen i Sverige bör tas hänsyn till. Studien ska också ta hänsyn till och inkludera relevanta delar från utvecklingen av kriterier för utvärdering inom området konflikthantering (ref. Sida Evaluation nr 00/37).

Studien ska översiktligt beröra erfarenheterna av statistiska attitydundersökningar (bl.a. Afrobarometer, World Values Survey).

Vad gäller diskursen på området ska den parlamentariska utredningen Globkoms skrivning om "Styrning, lärande och analys" inkluderas. Dessutom ska frågan om utvärderings- respektive controllerfunktionens perspektiv och inflytande på rapporteringen diskuteras, liksom konsekvenserna av att ta hänsyn till i vems intresse resultatrapportering sker. Eventuella konsekvenser av DESAs uppdrag att arbeta med ett rättighetsperspektiv i all verksamhet bör genomsyra diskussionen.

Förutom begrepp och metodfrågor identifierade av konsulten, bör bland annat följande begrepp och grundläggande metodfrågor kortfattat belysas med bl.a. referens till Sida/UTV manus till Utvärderingshandbok:

- resultatredovisningens olika syften;
- resultatredovisningens relation till utvärdering;
- resultatredovisning på aktivitets- respektive långsiktig effektnivå, liksom i förhållande till output och outcome;
- olika utvärderingsmodeller (måluppfyllelseutvärdering, intressent-utvärdering etc.);
- olika forskningstraditioner, varav främst positivism (naturvetenskaplig tradition) och hermeneutik (socialvetenskaplig tradition);
- olika typer av kriterier för bedömning av resultat, inklusive relevans och koherents;
- utvärdering och redovisning av kvalitet;
- djupintervjuer efter avsevärd tid efter en insats; samt
- definition av indikator.

6. Metod

För att genomföra uppdraget ska

- en litteraturstudie göras som ringar in de frågor som uppdraget ska belysa. Akademisk litteratur om resultatredovisning inklusive utvärderingslitteratur, liksom litteratur och rapporter från relevanta nationella och internationella biståndsaktörer inom området mänskliga rättigheter och demokrati ska utgöra huvudsaklig informationskälla för uppdraget.
- kompletterande intervjuer ska göras där särskilda behov föreligger. Häri kan inkluderas kontakter med etablerade akademiker inom relevanta områden. Uppdragstagarens kostnader för eventuella resor för detta syfte ska bestämmas i samråd med Sida/DESA.
- Sida/DESA ska utse en kontaktperson som i samråd med Sida /UTV vid behov ska vara dialogpartner under studiens gång.
- Uppdragstagaren ska efter 1 vecka till Sida presentera ett utkast till plan över innehåll och omfattning.

7. Tidsplanering

Studien beräknas ta 6 veckor (30 dagar) motsvarande heltid i anspråk av en person. Studien beräknas genomföras under juli – september månad 2002, och slutrapport lämnas senast 15 september, 2002.

8. Rapportering

Ett utkast till rapport ska presenteras till Sida/DESA 2 veckor innan sista dag för slutrapporten. Sida/DESA har då 5 arbetsdagar på sig att till uppdragstagaren komma med skriftliga kommentarer på utkastet för vidare bearbetning av uppdragstagaren.

Uppdraget ska resultera i en skriftlig slutrapport om maximalt 35 sidor, inklusive sammanfattning på

högst 3 sidor, men exklusive eventuella bilagor. Slutrapporten ska levereras till Sida/DESA senast den 15 september 2002 i elektronisk form i programmet Word, samt två exemplar i pappersform.

Uppdragstagaren skall efter att slutrapporten har lämnats stå till DESAs förfogande som resursperson under sammanlagt två dagar under andra halvåret 2002 för uppföljande diskussioner, dels internt inom Sida och dels tillsammans med externa aktörer.

Annex 2. Annotated bibliography

The purpose of this annex is to present a brief review of what other agencies are doing, and what experiences they have to share. When the field of evaluating democracy and human rights is surveyed, one can identify at least five different categories of publications that are of interest:

1. Handbooks and manuals on how to approach evaluation. Many of these are quite similar, but I will provide references and point to the highlights of a few of them.
2. Research studies that discuss democracy, human rights and development assistance, and that do so in a way that could be useful for those who do evaluations in the field.
3. Presentations of aggregate statistics, indicators of development and value surveys, that in various ways and various levels of ambition make global attempts to measure progress in democracy and human rights.
4. Synthesis reports, aggregate studies across donors and countries, that assess performance, lessons learned, and common experiences.
5. Individual evaluation reports, specific reports that I have come across that are of interest because of their methodological approaches, their ideas, or simply because they are good readings.

Handbooks and guidelines

As I mentioned above, several agencies have produced new evaluation manuals in the past few years. Prominent among these are Danida and Norad, and Sida is about to publish a new manual as well. Generally speaking, these manuals have limited practical applicability for an evaluator, as the primary audience is the staff members of the aid agencies. The manuals help them to define their own role in evaluation, to commission studies, and to become familiar with the most basic of methods and techniques to assess whether the job is reasonably well done. But for the evaluator who worries about how to measure impact, and who seeks solutions to how many interviews to conduct to get reliable data, they have little help to offer. Then there are manuals that provide detailed process guidance, as for example:

UNDP, Office of Evaluation and Strategic Planning, 1997. *Who are the Question-makers. A Participatory Evaluation Handbook*. 84 pages.

Available on-line at <http://www.undp.org/co/who.htm>

This is a handbook on how to conduct participatory evaluations. It provides an overview of literature in the field, and discusses the philosophy of participation. The most interesting part is the framework to define who is going to participate. The authors suggest that a matrix is used, which separates the steps in the evaluation process, and identifies the stakeholders. This can then be used to discuss which stakeholders ought to participate in different steps. It becomes possible to see who are included and who are excluded, and to target particular groups that one wants to see included. The handbook contains a case study module, which can be used for training in participatory evaluation. There are questions for group work as well as references to sources of expertise.

Fine, Allison H., Thayer, Colette E. and Cohan, Anne. 1998. *Program Evaluation Practice in the Nonprofit Sector*. Washington D.C. Innovation Network, Inc.

Although not exactly a handbook, this study aimed to determine which of two broad evaluation approaches – the traditional scientific model with an external evaluator collecting, interpreting and presenting data or the new participatory evaluation model – are being used, to what extent and with what effects. The research reveals that practice does not fall neatly into either category. The study focused on exploring the role of stakeholder participation in program evaluation. It examined whether different levels of stakeholder participation could be associated with organisational characteristics, evaluation characteristics, and evaluation outcomes. The findings offer useful guidelines for designing evaluation approaches emphasising stakeholder involvement. It is essential reading for anyone interested in participatory approaches to evaluation.

Fetterman, David M., Kaftarian, Shakeh J. and Wandersman, Abraham. Eds. 1996. *Empowerment Evaluation: Knowledge and Tools for Self-Assessment and Accountability*. Thousand Oaks, CA: Sage

This book represents a benchmark of sorts in that it attempts to stake out the territory of empowerment evaluation, and making explicit its ideological foundation. Fetterman's introductory chapter provides a comprehensive overview of the developing theory and practice in the field. The 15 other essays provide a good overview of current practice, as well as examples of tools, forms and checklists. Empowerment evaluation is the use of evaluation concepts, techniques and findings to foster improvement and self-determination. It is thus closely related to the subjects of democracy and good governance, in fact, it suggests a correlation between good governance and a particular, participatory, approach to evaluation.

OECD/DAC Working Party on Aid Evaluation. 1999 *Guidance for Evaluating Humanitarian Assistance in Complex Emergencies*. OECD, Paris.

Like many of the manuals produced by aid agencies, this one is also primarily aimed at those who commission and manage evaluations in donor organisations. I mention it here, because it is specifically geared to the one of subject of this paper, evaluation of human rights. It is divided into two parts. The first part describes humanitarian assistance programmes. It defines complex emergencies and outline the characteristics, opportunities and challenges of providing assistance in these areas. It describes the international response to emergencies; who the actors are, and how they roles differ (and overlap). One section describes the difference between humanitarian assistance and conventional aid programmes. The second part of the guidelines outline the evaluation process, and provides advice on how to write terms of reference, how to select a team, methods of working, and follow -up. It is available onn-line for organisations attached to the OECD database.

Jackson, Edward T. and Kassam, Yussuf, Eds, 1998. *Knowledge Shared: Participatory Evaluation in Development Cooperation*. Connecticut, USA: Kumarian Press and Ottawa Canada: IDRC.

The book consists of 13 papers into two parts: the first presents 4 discussion papers dealing with issues, strategies, and methods of participatory evaluation, and the second provides 9 case studies. It offers a good introduction to the topic. Chapter 3 contains a matrix illustrating how participatory evaluation and results-based evaluation could be brought together, showing outputs, outcomes and impacts by level of intervention. An in-depth and critical review of this book can be found on the website <http://www.capacity.org/pubs/annotations/issue2jackson.htm>

Nagel Stuart, S., Ed. 2002. *Handbook of Public Policy Evaluation*. Sage, London.

This newly published handbook presents a variety approaches to systematic policy evaluation. It describes theoretical ideas as well as practical applications, as for example win-win policy evaluation. Policy evaluation can distinguish between evaluation of the policy statement as such, whether it is clear,

consistent, coherent and relevant in light of the challenges in the policy domain. Furthermore, policy evaluation can be about the implementation of policy, how different policy instruments are used to realise the policy. There is also the question of effects, results and impact. Did the policy achieve its objectives and did it produce any unintended side effects? As policy evaluation is a complex set of hierarchical categories, there is a need for the type of enlightenment brought by this handbook.

Miller, Delbert C. and Salkind, Neil J., 2002. *Handbook of Research Design and Social Measurement* (6th edition), Sage; London.

This comprehensive manual (780 pages) answers most of the practical questions around sampling, survey techniques, interviewing, the practical construction of scales and measurements, and statistical techniques that those who actually do evaluations have. If, by chance, the answers are not in the book, then it contains a large number of references to other sources of information. This is the 6th edition of a classical volume, and one of the novelties is the number of references to web resources in evaluation and research design.

Patton, Michael Q. 2002. *Qualitative Research and Evaluation Methods*. Sage, London.

Michael Patton is a prolific writer on evaluation manuals. His “Utilization-focused Evaluation is a classical study, and so is this, which now appears in the 3rd edition. It is a useful supplement to the above mentioned text, which is quantitatively oriented. This book explores the practices of qualitative research, for example different interview techniques, focus groups, story telling etc. It explores how to make use of the web in participatory and qualitative research settings. It is full of lively examples, and a very good source of inspiration.

The World Bank. 2001. *Evaluating Public Sector reform Programs. Guidelines for Assessing Country – level impact of Structural Reform and Capacity Building in the Public Sector*. Operations Evaluation Department, The World Bank, Washington D.C.

Even though more than half of this rather brief manual describe the subject and accounts for World Bank policies and experience in the field of public sector reform, the remaining half is a useful and practical guide to assessing country level impact. The assessment is structured in two dimensions; (1) responsiveness and relevance, and (2) efficacy and impact. What I like most in the book is its illuminating use of models to organise the inquiry. In practice, the indicators would probably have to be tailor-made more to the task, but the suggestions here facilitate that work. In the annexes, there are further refined examples of indicators, interview formats and survey instruments that can be adapted. All in all, it is very useful guide, and it is aimed at those who do the job evaluation, and it is for them that it is practically useful.

Research; books, papers and conference proceedings.

It is of course impossible to have a total overview of relevant books, papers, and other publications. I provide some few examples that from the fields of political science and evaluation. I apologize because it is very incomplete and I am sure I have missed a number of more relevant and interesting publications. Hopefully those who read the studies mentioned here would, in them, find references that lead them onwards.

Bemelmans-Videc, Marie-Louise, Rist, Ray, C. And Vedung, Evert, Ed. 1998. *Carrots, Sticks and Sermons: Policy Instruments and Their Evaluation*. Transaction Publishers, New Brunswick and London.

This study offers a comprehensive analysis of categories and typologies of policy instruments. It classifies sticks, carrots and sermons – or, more specifically, regulation, economic means, and information.

There are comparative perspectives on evaluation in different environments, and also guidelines and insights on the differences in evaluating the three categories of policy instruments. The book contains the models that I mentioned in chapter 5, and is a good source of inspiration for anyone who will be engaged in policy evaluation. The book builds on experiences from diverse settings, such as Belgium, England, Korea, Sweden, the Netherlands, Canada, Korea and the United States.

Beck, Ulrich, 1995 (in German 1993) *Att uppfinna det politiska. Bidrag till en teori om reflexiv modernisering*. Daidalos, Göteborg.

This is a classical study in political science, discussing the role of politics in modern societies perhaps at present, but even more so in the future. The author puts question marks to rational modes of governance and political decision-making, and introduces new thoughts about the need for politics in the post-modern era. The book discusses ambivalence, uncertainty and ambiguity as central aspects of life, and explores the role of politics in embracing the ensuing complexities. It is perhaps an unconventional approach in the context of development cooperation, but I do not see any point in refraining from the difficult, challenging, and provocative ideas – on the contrary, actors in this field may accept the complexities that such perspectives bring to their subject.

Leduc, Lawrence, Niemi Richard, G., and Norris, Pippa. Ed., 2002. *Comparing Democracies 2*. Sage, London.

The aim of this book is to introduce and understand representative democracy as a political process and contemporary systems of governance in need of constant attention and scrutiny. The book consists of 10 essays, where the first chapter deals with general comparative methodology. The remaining chapter discuss different aspects of democratic governance, such as, electoral systems, referendums and other aspects of direct democracy, party systems, candidate selection, campaign communications, political cleavages, issues and electoral change. The final essay discusses consolidating democracies. The study is a fundamental companion to a detailed evaluation of governance issues, and even though it is no manual, the practical examples of how the studies were undertaken are enlightening. There are many useful references.

Kooiman, Jan, 2002. *Governing as Governance*. Sage, London.

As we all know who occasionally work with development cooperation, the concept of governance has become central catchword for everything that should be improved and done better. This book maps the field and demonstrates the utility of a socio-political perspective to understand contemporary forms of governing, governance and governability. The central underlying theme of the book is the notion of governance as a process of interaction between different social and political actors and the growing interdependencies between the two as modern societies become ever more complex, dynamic and diverse.

Thomas, Alan, Chataway, Joanna, and Wuyts, Marc. 1998. *Finding out Fast; Investigative Skills for Policy and Development*. Sage, London.

Recognising that policy decisions are typically made under pressure of time and on the basis of incomplete data or with limited resources with which to obtain information, the authors provide information on how to locate, evaluate and use relevant information fast. It is a very practical collection of essays. The best ones deal with the analysis of data, and how you draw conclusions. There is also a very good chapter on how evaluators should write their reports so as to grab the attention of the readers. A much needed and very difficult subject! Even though the authors discuss development cooperation in general, their book is very relevant for the subject of this paper.

Dahlberg, M. and Vedung, E., 2001. Demokrati och brukarutvärdering. Studentlitteratur, Lund.

The authors explore the potential and the limits for participatory evaluation. Their chief subject is to assess participatory evaluation, and thus by extension other ways and means of engaging the public more directly in policy formulation as well as implementation of policies. Does this stand in contrast to traditional models of representative democracy? What happens if the most vocal interest group are allowed to influence policy formulation and resource allocation above and beyond the influence that they – as well as all other citizens – have in the election of representatives on election days. If the users of a particular service are to have more influence other than this, which organisational means stand at the disposal of the political system? These are questions that need to be asked and debated seriously, and much of the literature on participation fail to do in their enthusiasm over the participatory “ideology”. The book then analyses an example of participatory evaluation in the social services.

Aggregate statistics, indicators and value surveys,

When I reviewed Robert Putnam’s studies of democracy in Italy and the U.S., it was evident that the research work was facilitated by the enormous amount of statistical data that was to be had, for example in the form of value surveys, opinion polls, and also the regular annual statistics on habits, consumption, and production that are produced by the national statistical agencies. There are now a number of organisations that provide such data on developing countries too, and that focus on values and attitudes, as well as “hard facts” relating to the progress of democracy and the protection of human rights.

Afrobarometer.

The Afrobarometer is a survey research instrument that measures the social, political and economic atmosphere in Africa. It is meant to reveal what Africans think about democracy and development, about their own roles in these processes and various aspects of government performance. It is meant to help empower African citizens. The Afrobarometer is produced by partners in different countries. The partners produce national reports, and these in turn are to be aggregated to regional reports. There will be outreach seminars to different audiences. The project is in its starting phase, so I have not seen any comprehensive reports yet. There are national datasets available at www.icpsr.org and www.idasa.org where it is possible to look both at the survey instrument and the first reports.

Amnesty International, Annual report and country surveys. www.amnesty.com

The reviews made by Amnesty are well researched and down to earth. They report on the lack of democracy and respect of human rights, rather than on its presence and progress. Obviously, a report that described the malpractice in political culture is as relevant as one that records progress. In addition, if one wants to trace change, then a comparison over time of Amnesty reports is as relevant as any other measure. One question remains, the Amnesty report are qualitative surveys, so should they be seen as indicators? Yes, definitely, the definition of what an indicator is does emphasise that indicators can be qualitative. A descriptive text of 20 pages, can be as much of an indicator as number in a table. Furthermore, the qualitative text of the Amnesty reports do make use of statistics at a “lower level”, as for example the number of political prisoners, refugees, cases brought to international courts, etc. But these statistics are put in a context and used for a comprehensive assessment of situations.

Civics Civil Society Index, Civics New Civic Atlas, 1997, and www.civics.org

The civil society index is an assessment tool for practitioners and policy makers. It offers a methodology for generating information about the well-being of civil society in any country or region, including its strengths, challenges and shortcomings. The components of the index are: (1) structure, that is, how large civil society is and its component parts, (2) value, that is, the measuring the underlying values of

civil society, (3) space, that is, the legal, political and socio-cultural space in which civil society operates, and (4) impact, that is, civil society's contribution to social, economic and political problem-solving. Some of these measures are highly problematic, as there is bound to be diverse values in different civil society organisations, (in Sweden, compare the scouting movement, to football leagues, and jazz clubs to see the difficulties entailed in a single measure of values).

The United Nations Development Programme (UNDP). Human Development Report. Oxford University Press, Oxford.

The Human Development Report of 2002 focuses on democracy; it has the subtitle "Deepening democracy in a fragmented world. Apart from the various statistics on human development, this report puts forward a number of indicators on democracy and governance. The UNDP actually tried to develop a democracy index in previous versions of the annual human development report (years 1990 and 1991), but these were not continued, presumably because they met opposition from some member states. It remains to be seen whether the measures presented in the report of 2002 can be continued. The measures used in the report are not compiled by UNDP itself, but accessed through other data sets, for example the World Bank Governance Indicators, Transparency International, Freedom House and others. The UNDP report puts data from many sources together, and thus presents a fuller picture than do any of the quoted sources in isolation. In addition, after the indicators and the measures are presented, there follows some 100 pages of analytical text dealing with the relation between democracy, economic development, and poverty. The indicators are interesting, but it is even more interesting how the figures are put to use in the analytical sections of the report.

The World Bank. World Development Report. Oxford University Press, Oxford

The annual reports from the World Bank also have different focus from one year to the other. Over the past decade, they have focused on poverty, governance, environment, etc. As the issues of human rights and democracy are of an interdisciplinary nature, there is always something of interest and relevance in the reports. The World Development Report furthermore has a number of statistical tables, many of which contain indicators of aspects of development that relate to good governance. There are, for example statistics on women in development, on income distribution, on urban pressure, etc. Compared to the UNDP report, the focus in the Bank's report is more on the economic side of development, but there are of course a number of such statistics that are necessary to obtain in order to have a proper understanding of the broader framework in which progress – or lack of progress – in relation to the more direct measures of democracy take place. I think particularly of statistics in respect of government expenditure, public debt, and other balance of payments related figures. The World Bank report is a necessary supplement to the Human development Report, and vice versa.

Synthesis reports, studies across donors and countries, that assess performance, lessons learned, and common experiences.

In many aid agencies, the line departments are responsible for conducting evaluations of specific projects and programmes. Central evaluation units, on the other hand, often see as their role to produce synthesis studies, to analyse cross-cutting themes, country and sector analysis. International organisations at times also see their specific advantage as one of doing studies of cross-cutting themes, then also comparing the experiences of different donors in a field. Over the past couple of years, several studies of this kind were produced in the field of democracy and human rights. Again, I cannot but apologise for a rather random selection of such studies, and I may have missed more recent contributions. What I have used, is the recent survey by Crawford and Kearton (2002), which is also summarised below, and my assessment of the Danish and OECD reports builds on some of comments made by them (theirs is a fuller and more comprehensive analysis).

Crawford, Gordon, and Kearton, Iain. 2002. *Evaluating Democracy and Governance Assistance*. Centre for Development Studies, University of Leeds.

The authors have contacted aid agencies in order to build a database of evaluations. They gathered 110 evaluations, which they catalogued. Their paper explored the experiences of seven donor agencies (CIDA, Danida, Sida, USAID, UNDP, The European Commission and the OECD – the latter is of course not an aid agency, but has been active in evaluation). The approaches taken differ, but what most have in common are a fairly hesitant approach to evaluation. USAID is the one different agency, as it has whole-heartedly gone in for a quantitative approach with the help of a set of indicators of country progress in terms of democracy. Rather, I should say that the theory is clear and the structure of the report is clear. It is less clear whether it works in practice and whether the results are valid enough so that decision-makers dare use them. Following the review of donor experience, the authors proceed to discuss challenges in evaluating democracy, and they advance an alternative methodology (alternative to the USAID approach, and alternative to a view of impact as something to be discussed in logframe terms). The suggestions for a developed methodology focus on; (1) political context analysis, (2) meso level analysis, and (3) participatory methods. The paper ends with a continued discussion of evaluation reports. It is the most comprehensive review to date, and the practical recommendations are down to earth and realistic, and their critique of donor practice appears well-founded.

Danida. 2000. *Evaluation of Danish Support to Promotion of Human Rights and Democratisation 1990 – 1998*. Ministry of Foreign Affairs, Copenhagen.

This large scale evaluation was undertaken in the late 1990s. The final reports consist of 9 volumes; a synthesis volume, four thematic studies, and four country studies. It was undertaken first and foremost as a learning exercise. It is described as “self-critical learning, rather than accountability”. However, it is quite clear that an informed and penetrating discussion of accountability is much furthered – indeed impossible without – a study such as this. It lies in the nature of the task (compare the Swedish parliamentary inquiries reviewed in chapter 2) that the study raises more questions than it answers. It raises fundamental and challenging issues, notably concerning methodology and impact evaluation – and at times it may seem to provide fewer answers than are actually possible. Nevertheless, the study represents a considerable achievement, and it is essential reading.

OECD. 1997. *Evaluation of Programmes Promoting Participatory Development and Good Governance: Synthesis Report*. OECD, Paris.

This report summarises five evaluations (which in turn also were synthesis reports) conducted by member countries. The agencies divided labour according to various themes; so the U.S. contributed with an evaluation of assistance to legal systems, the UK with a study of public sector management, the Netherlands with an evaluation of human rights assistance, Norway with a study of decentralisation, and Sweden with one of participation. There are no common methods between the thematic papers, far less methodology. The overall objective of the study was to learn lessons from donor experience, and these were outlined in the introductory chapter, but also specified in the reports of each theme. It is hard to assess how valuable such an exercise is. The studies build on synthesis of already existing project and program evaluations. Hence, the conclusions of the synthesis cannot be more interesting than these studies were. Many of the lessons learned appear rather trivial, as for example:

- Reform efforts require political commitment
- Donor advocacy of “home-grown” initiatives are likely to be more successful than donor-driven reform efforts
- Participation and participatory approaches should be important ingredients....

..... and so on. It is hard to deny the truth of these propositions, but did anyone believe the contrary in the first place? Who is to be convinced? Is there a risk that lessons learned at that level of abstraction, after so many synthesis reports, almost become meaningless? As I am not an expert in the field, I probably fail to see the significance of the lessons learned, but I cannot refrain from asking the reader to reflect on what is new, what is well-known, and what is worth analysing in another study.

There are many specific evaluation to analyse. However, a selection here would be very random, and I would suggest that the readers of this report consult the databases of the aid agencies that of particular interest, or turn to the synthesis reports as well as to the handbooks and manuals mentioned above. These contain good references to other evaluations that are worth studying in detail.



SWEDISH INTERNATIONAL DEVELOPMENT COOPERATION AGENCY
S-105 25 Stockholm, Sweden
Tel: +46 (0)8-698 50 00. Fax: +46 (0)8-20 88 64
Homepage: <http://www.sida.se>