

Extract from the Commission's Evalsed

4) Managing quality assurance and quality control

Assessing the quality of an evaluation is an integral and fundamental part of the evaluation process. An evaluation that does not meet some minimum quality standards could mislead decision-makers and programme managers.

However, to assess evaluation quality is a complex process. The evaluations performed in the context of socio-economic development programmes and policies are too different from each other to allow the existence of few simple rules that can guarantee the quality across the board. By and large one can say that the quality of the evaluation as a whole is conditional upon the presence of three distinct but interrelated aspects:

- the quality of the planning and design phase, including the commissioning of the evaluation;
- the quality of the implementation of the evaluation itself; ▪
the quality of the available data.

These aspects are interrelated in the sense that poor performance by the evaluator can very well stem from the poor quality of the data or from the flaws of the planning and design phase. Unfortunately those involved in these three sets of activities are different and often their goals, as well as their quality criteria, are also different. For instance the monitoring system designed for the day to day management of the programme does not necessarily produce the data needed for an evaluation of impacts.

In the first place, quality can be considered a *characteristic of the process* through which the evaluation activities are performed. The assessment of quality could include: the way in which the commissioning authority develops the decision to proceed to an evaluation, defines its scope and the resources available. This can be analysed in order to understand if the procedures followed were appropriate to the allocation of the different responsibilities, if the contribution of the various stakeholders was taken into consideration, etc.. The same goes for the performance of the evaluation. One can focus on the way in which the team, and its interaction with the commissioner and the evaluators, was managed, the checks that were put in place in order to ensure that the data collected were properly treated, etc.. The organisation of the monitoring process can be assessed as well.

In the second place, quality is a *characteristic of the products* of the evaluation process. Thus one could analyse the ToR according to the criteria that we have already spelled out. Of course, one can assess the quality of the intermediate and final evaluation reports to see whether they meet some basic criteria of good professional practice and if the data are sufficient in quantity and reliable enough to warrant sound judgements.

In theory the two aspects the process and the product are linked: a good process should generate a good product.

It is now becoming common to define good practice standards in evaluation. These have been elaborated by international bodies (such as the OECD), National Administrations (for example, the Italian Department for Economics and Finance) or professional associations such as national evaluation societies and associations. Many of these follow on from earlier efforts in the United States and can be traced back to American Evaluation Association (AEA): *Guiding Principles for Evaluators* (1992) and the Joint Committee on Standards for Educational Evaluation Program *Evaluation Standards* (1994).

Standards Guidelines and Ethical Codes:

- USA - Program Evaluation Standards (1994) Joint Committee on Standards for Educational Evaluation, Program Evaluation Standards: <http://www.jcsee.org/program-evaluationstandards>
- Deutsche Gesellschaft fuer Evaluation (DeGEval): Standards fuer Evaluation (2001) : <http://www.degeval.de/degeval-standards>
- Switzerland: SEVAL Evaluation Standards: http://www.seval.ch/de/documents/seval_Standards_2001_dt.pdf
- The African Evaluation Guidelines 2000: https://www.globalhivmeinfo.org/Gamet/Gamet%20Library/1295_African%20Evaluation%20Guidelines.pdf
- American Evaluation Association (AEA), Guiding Principles for Evaluators : http://www.eval.org/GPTraining/GP%20Training%20Final/gp_principles.pdf
- Australasian Evaluation Society (AES), Guidelines for the Ethical Conduct of Evaluations: http://www.aes.asn.au/about/Documents%20-%20ongoing/guidelines_for_the_ethical_conduct_of_evaluations.pdf
- UK Guidelines for good practice: <http://www.evaluation.org.uk/about-us/publications>
- PUMA Best Practice Guidelines : <http://www.oecd.org/dataoecd/11/56/1902965.pdf>
- Italy Treasury Guidelines: http://www.dps.tesoro.it/documentazione/docs/all/Criteri_qualita_sistema_nazionale_valutazione_maggio2002.pdf

This list provides a cross section of some current evaluation standards and codes. Most, in particular those that derive from the AEA Joint Standards such as the German Evaluation Society's (DeGEval) and the African Evaluation Guidelines are directed primarily at the technical conduct of evaluation by evaluators. The Australasian and to some extent the UK Evaluation Society's outputs are more concerned with ethical codes of practice. More recently a further category of guideline has emerged. This is directed more at those who commission evaluations. Examples of this can be found in the OECD (PUMA and DAC guidelines) and in the European Commission.

Although there is not yet consensus about all the components of a quality assurance system for evaluation, we see a shift from a focus largely on quality control, i.e., ways of judging report/ output quality. This shift was endorsed by a number of years ago by study on the use of evaluation by the European Commission.

Quotation from EU research on use of evaluation:

[The study] ... tends to support the value of inclusive standards that encompass the interests of commissioners of evaluation, evaluators and citizens. Broader European evaluation standards (instrumental and ethical) as are being considered by European Evaluation Society and several other European national evaluation societies could complement the move towards standards developed by the European Commission and some National Administrations (The Use of Evaluation in Commission Services October 2002).

Both quality control and quality assurance criteria are suggested below. Normally the person responsible for managing the evaluation within the commissioning body takes responsibility for applying the quality control criteria. Performance on the quality assurance criteria should be informed by the views of members of the Steering Committee, other stakeholders, the evaluation team and those responsible for managing the evaluation in the commissioning body. The Steering Committee should provide the criteria as early as possible in the evaluation assignment and is normally best placed to make the overall assessment at the completion of the work. For quality control, consultation with external experts or referees can be useful. The application of quality control/content-type criteria and quality assurance/process-type criteria are undertaken for different purposes. Quality control of content offers assurance that the work has been properly conducted and that its conclusions can be relied on. Quality assurance of the evaluation process will contribute more to learning about evaluation management and provide inputs that should improve future evaluation management.

Quality control and quality assurance criteria

Quality Control: Output Criteria	Quality Assurance: Process Criteria
Meeting needs as laid out in ToR	Coherent and evaluable objectives
Relevant scope and coverage	Well drawn terms of reference
Defensible design and methods	Sound tender selection process
Reliable data used	Effective dialogue and feedback throughout evaluation process
Sound analysis	Adequate information resources available
Credible results that relate to analysis and data	Good management and co-ordination by evaluation team
Impartial conclusions showing no bias and demonstrating sound judgement	Effective dissemination of reports/outputs to Steering Committee and policy/programme managers
Clear report with executive summaries and annexed supportive data	Effective dissemination to stakeholders

Quality control - output criteria

▪ Meeting needs

Has the evaluation answered the questions of the ToR satisfactorily and does the report provide additional information that might be essential for the commissioners? In particular:

- Has the way programme or intervention objectives evolved and been analysed?
- Does the report cover the entire programme? If not, is the selection justified as regards the priorities stated by the commissioners in the ToR and subsequently?
- Does the evaluation provide useful feedback for programme managers? Does it include lessons on successes and failures that may be of interest to other programmes, regions or countries?

▪ Relevant scope

Have the essential characteristics of the programme or intervention have been described and have problems and successes in implementation have been properly clarified?

Has the evaluation overlooked other potential or future results, as well as any unexpected yet significant effects and results?

Finally, it is necessary to check whether:

- the limits of the scope, in terms of areas or groups, are defined according to the logic of the intervention;
- the scope includes peripheral areas or non-eligible groups which are nevertheless likely to be affected by the evaluated interventions;
- the evaluation considers the evaluated programme or intervention in isolation or includes its interactions with other European or national programmes.

▪ Defensible design

Methodological choices must be derived from the evaluation questions. The evaluation must, moreover, make the best possible use of existing research and analyses:

- Has relevant knowledge been collected and used wisely?
- Is the method or the combination of methods justified for answering the evaluation questions properly?
- Were the reference situations chosen (counterfactual or other) appropriate for comparisons?

Any evaluation report must include a description of the methodology used and clearly define the sources of data. Similarly, the limits of the combination of methods and the tools used must be clearly described. It is necessary to check whether:

- the methodology is described in enough detail for the quality to be judged;
- the validity of data collected is clearly indicated;
- the available data correspond to the methods used.

Finally, does the evaluation adequately analyse relations of cause and effect for the most essential questions?

▪ Reliable data

Evaluators use secondary data and primary data they have collected for the evaluation. In the latter case, the methods used to collect and process the data are important factors in the reliability and validity of the results:

- Are available sources of information identified and the reliability of this data checked?
- Have data from the monitoring system and previous studies been used optimally?
- Were techniques used to collect data complete and suitable for answering evaluation questions?

Whether the collection of data used quantitative or qualitative techniques or a combination, it is necessary to inquire if:

- the mixture of qualitative and quantitative data is appropriate for valid analysis;
- the "populations" used for data collection have been correctly defined;

- the survey samples or cases studied have been selected in relation to established criteria.

▪ Sound analysis

Quantitative analysis consists of the systematic analysis of data using statistical and other techniques. It has a particular focus on numerical values. Qualitative analysis consists of the systematic comparison and interpretation of information sources with a particular focus on why things happen. In both cases it is necessary to assess whether the methods of analysis used are relevant to the data collected and whether the analysis has been carried out to an appropriate quality. Relations of cause and effect are complex and therefore constitute a particular challenge for evaluation. It is necessary to check:

- whether the relations of cause and effect underlying the programme are explicit and relevant so that the object of analysis can be focused, and
- to what extent the analysis uses suitable techniques.

For this reason, a comparison between beneficiaries and a control group or at least a before-after comparison, is recommended, with an explanation for the choice of approach given.

▪ Credible results

The credibility of results requires that they follow logically and are justified by the analysis of data and interpretations based on clearly presented explanatory hypotheses. The validity of the results means that the balance between internal validity (absence of technical bias in the collection and processing of data) and external validity (representativeness of results) must be justifiable. The need to perform in-depth analyses of a part of the programme poses the problem of extrapolation, from case studies, for the programme as a whole. In this context, it is necessary to check that:

- the interpretative hypotheses and extrapolations are justifiable and the limits of validity have been defined;
- the selection of cases and samples makes it possible to generalise the findings.

▪ Impartial conclusions

Conclusions include suggestions and sometimes recommendations. Whereas results are "technical" and can be analysed without too much risk of impartiality, conclusions and recommendations are issued on the basis of value judgements. The quality of the judgement is thus decisive. To answer the question: Are the conclusions fair, free of personal or partisan considerations and detailed enough to be implemented concretely, it is necessary to check that:

- the elements on which the conclusions are based are clear;
- the conclusions and recommendations are operational and sufficiently explicit to be implemented;
- controversial questions are presented in a fair and balanced way.

The evaluation questions of the ToR must be answered.

▪ Clear report

Evaluation results can be disseminated and communicated to the stakeholders in writing or verbally. The final report is only one means of diffusion and continual communication of results is desirable. The

clarity of the report will depend on the quality of the presentation of results and the limits of the work performed. It is necessary to check that:

- the report was written clearly and is set out logically;
- specialised concepts are used only when necessary and they are clearly defined;
- presentation, tables and graphs enhance the legibility and intelligibility of the report; and
- the limits of the evaluation, in terms of scope, methods and conclusions, are clearly shown.

In many cases only the summary of a report is read so this must be clear and concise. It must present the main conclusions and recommendations in a balanced and impartial manner. It must be easy to read without the need to refer to the rest of the report. ■ **Quality assurance criteria**

The next set of criteria concerns the overall process and context of the evaluation: quality assurance rather than quality control. It will allow those assessing quality both to understand what might account for positive and negative aspects of the evaluation and draw lessons that could be applied in order to improve the quality of future evaluations.

- Coherent and evaluable objectives

The coherence of the objectives: the extent to which they are specific, linked to interventions, not contradictory, etc., has been discussed earlier. It was noted that the use of logic models, programme theory and theory of change approaches are useful ways to clarify programme objectives and the logic of interventions at the early stages of a programme prior to the launch of an evaluation. At this stage we are interested in the outcomes of this earlier process. How far were the evaluators dealing with a coherent programme or intervention in terms of objectives? Were any evaluation difficulties the result of poorly articulated objectives or other problems of evaluability?

- Well drawn terms of reference

Sound terms of reference make for effective evaluations. It is possible at the time they are drafted to judge the adequacy of a ToR. It becomes easier with hindsight to identify what might have usefully been included. This is important for future learning. A poor or incomplete ToR can lead evaluators to deploy their resources inappropriately.

- Sound tender selection process

Was the tender selection process well conducted? This is both a procedural question and a matter of substance. Procedurally an assessment should be made of the systematic application of relevant criteria at selection. Substantively we are interested in whether the right decision was made.

- Effective dialogue and feedback throughout evaluation process

Keeping an evaluation on track, providing feedback and providing a forum for stakeholders to learn through dialogue with each other and with the evaluators is a recognised prerequisite for quality. This is partly a question of the forum created for this purpose, the Steering Committee, but possibly also specific briefing meetings and workshops. The inclusiveness of the membership of such meeting places needs to be assessed: were the right stakeholders involved?

Was good use made of Steering Committee meetings? Were the agendas appropriate? Did stakeholders see these opportunities as productive and enhancing their understandings? Did they ultimately help shape and improve the quality and usefulness of the evaluation?

- Adequate information resources available

Evaluators need information. An assessment needs to be made of the adequacy of information. Monitoring systems should help programme managers and an evaluation usually reveals the extent to which they do. Evaluators will also need to draw on secondary administrative data, gathered often for other purposes by local, regional and national administrations.

Much information in an evaluation is held in the minds of key informants. This is especially so for contextual and qualitative information which is important not only to understand the intervention but also how to interpret more formal data.

To judge the quality of the process and context of the evaluation there needs to be an assessment first of whether information existed and second whether it was made available. Judgements about the availability of information and data to evaluators can provide data about the actual state of partnership and inter-agency cooperation.

- Good management and co-ordination by evaluation team

However well planned and however convincing the work plan and inception report, all evaluations need to be executed properly. They need both to follow plans and be able to adapt to unexpected events that make plans - or aspects of them - redundant. Teams need to be kept together and the different work components need to be co-ordinated and outputs integrated. Relations with commissioners of evaluation, programme managers and a variety of informants, implicated institutions, groups and associations have to be managed. These aspects of management are mainly the responsibility of the evaluation team and its managers. However, some elements are shared with programme managers and those responsible for commissioning the evaluation.

- Effective dissemination of reports/outputs to Steering Committee and policy/programme managers

Report dissemination is another shared responsibility. In part it depends on the ability of the evaluation team to produce high quality and well-drafted outputs. It also requires an awareness of the value and opportunities for dissemination within the evaluation team. There is a big difference between evaluators who confine their feedback to the contractual minimum and those who provide ad hoc feedback when new problems occur or when key issues need to be resolved.

Dissemination also requires sensitivity to the information needs and interests of key stakeholders. Sometimes outputs need to be tailored to meet different interests.

- Effective dissemination to stakeholders

An evaluation process should not be considered complete until a programme of dissemination has taken place. The general requirements for such dissemination should have been signalled in the ToR. Primary responsibility does not rest with evaluators. Programme managers and those who commission evaluations should take responsibility for dissemination, including to the public at large. ■ **The synthetic assessment**

A synthetic assessment recapitulates the quality criteria. It is difficult to recommend any particular weighting for the different criteria because their importance varies from one situation to the next. A five point rating scale is used. Thus there are two positive possibilities and two negative possibilities and a mid-point when the balance of judgement is uncertain.

<i>Please assess the evaluation report in terms of your judgements as to how positively or negatively it met each criterion specified below:</i>	Very positive	Very negative
---	----------------------	----------------------

<p>1. Meeting needs: The evaluation report adequately addresses the requests for information formulated by the commissioners and corresponds to the terms of reference</p>	<input type="checkbox"/>				
<p>2. Relevant scope: The rationale of the programme, its outputs, results, impacts, interactions with other policies and unexpected effects have been carefully studied</p>	<input type="checkbox"/>				
<p>3. Open process: The interested parties – both the partners of the programme and the other stakeholders – have been involved in the design of the evaluation and in the discussion of the results in order to take into account their different points of view</p>	<input type="checkbox"/>				
<p>4. Defensible design: The design of the evaluation was appropriate and adequate for obtaining the results (within their limits of validity) needed to answer the main evaluative questions</p>	<input type="checkbox"/>				
<p>5. Reliable data: The primary and secondary data collected or selected are suitable and reliable in terms of the expected use</p>	<input type="checkbox"/>				
<p>6. Sound analysis: Quantitative and qualitative data were analysed in accordance with established conventions, and in ways appropriate to answer the evaluation questions correctly</p>	<input type="checkbox"/>				
<p>7. Credible results: The results are logical and justified by the analysis of data and by suitable interpretations and hypotheses</p>	<input type="checkbox"/>				
<p>8. Impartial conclusions: The conclusions are justified and unbiased</p>	<input type="checkbox"/>				
<p>9. Clear report: The report describes the context and goal, as well as the organisation and results of the programme in such a way that the information provided is easily understood</p>	<input type="checkbox"/>				
<p>10. Useful recommendations: The report provides recommendations that are detailed enough to be implemented</p>	<input type="checkbox"/>				
<p>In view of the contextual constraints bearing on the evaluation, the evaluation report is considered to be:</p>	<input type="checkbox"/>				

The use of evaluation and knowledge management

Undertaking evaluation and ensuring its quality is only worthwhile if the activity leads to some use of the evaluation findings and contributes to improved knowledge amongst those best able to take advantage from it. There are at least three different ways in which evaluation is used.

- Evaluations may be used directly in an instrumental manner when the results, findings, conclusions and recommendations are taken up. In practice this is unusual and where it does occur it tends to take place only partially.
- More often, several evaluations or evaluations combined with other evidence and opinion are used cumulatively to inform debates and influence decision-making. Evaluation thus stimulates the process of debate, challenge and counter challenge to evidence and its interpretation.
- Even where evaluation results are not used the process of evaluation initiation and reflection can be useful by offering opportunities to exchange information and clarify thinking.

The extent of use of evaluation and its impact is influenced by a number of factors:

- The organisational arrangements for dissemination. The time and resources available for dissemination and the degree to which the process is championed by those responsible for the work influences the understanding, communication and use of the findings.
- The quality of the evaluation. Where evaluation standards are high the results cannot be easily dismissed.
- The involvement of stakeholders in the stages of the evaluation cycle alongside evaluators and administrators. This is essential to build up evaluation use.
- The involvement of senior managers. This helps ensure that policy and resource allocation as well as practice are influenced by evaluation findings.
- The application of a system of systematic follow up of the conclusions of evaluation. This process both draws attention to where the findings have been and have not been used and reduces the tendency to relearn the same lesson.
- The institutional arrangements for conducting evaluation. There are no perfect models. Evaluation findings are likely to be of use to decision makers, those involved in the planning and design of interventions, and those involved operationally. The tendency towards the organizational separation of evaluation, operational and policy functions may lead to the improved independence and quality of evaluation. Policy and operational concerns can for example over emphasize what can be achieved through evaluation. On the other hand the separation but may be less helpful if it leads to an overemphasis on evaluation management and limits the use of the evaluation.

It is reasonable to conclude that the creation of an evaluation culture is essential for organisational learning. Key components of an evaluation culture over and above the generation of quality evaluations include: a presumption that interventions should be designed and implemented in a manner that facilitates subsequent evaluation; an appreciation of the range of purposes of evaluation; a recognition of the limits of evaluation, the scope for interpretation and the need to combine quantitative and qualitative evidence; and a recognition of the needs of different users of evaluation.

Golden rules

1. Evaluation competence should be brought in early by planners. This can help to clarify objectives. This activity, although employing evaluation competence, is quite separate from mainstream evaluation activities. It needs to occur at the design and planning stage. However, this can make subsequent evaluation easier and more successful. Evaluability assessment and preparing an analysis of programme theory can be used.
2. A similar evaluability assessment should be undertaken by evaluators when they begin their work. This may overlap or repeat what has already taken place at planning stage. However, the purpose is different. It is to ensure that a feasible evaluation plan is produced and to clarify how evaluation outputs will be used. This is consistent with a general expectation of evaluators that they should be concerned with how their results, conclusions and recommendations are used from the earliest possible stage of their work.
3. Stakeholders, programme managers and policy makers, potential beneficiaries and partners should be involved in the evaluation from the earliest stages, where practicable. This will ensure that the evaluation design and plan will include their priorities. It will also ensure that they feel some sense of ownership of the outputs of the evaluation and are more likely to find them useful and use them. It may, however, be necessary to be selective in deciding on which voices finally determine the evaluation agenda, to retain focus and ensure the evaluation is manageable. Overarching priorities should be shaped by the intentions and logic of the programme or intervention whilst remaining open to unintended consequences especially for intended beneficiaries.
4. Evaluations need to be actively but sensitively managed. This will ensure that commissioners are aware of choices to be made along the way and that evaluators receive sufficient support, access to information and briefing as to changes in policy and context. Those responsible for commissioning an evaluation and programme managers are the most suitable people to manage the evaluation because they are aware of its background and rationale.
5. It is usual to derive criteria for an evaluation, i.e., judgements as to the basis for positive and negative assessments of progress, from the objectives of a programme. It is also important to include a wider set of criteria that derive from social needs. For example, is this programme useful for those whom it is intended? Is the programme consistent with other policy initiatives? Maintaining this broader perspective ensures that for part of their work evaluators can stand outside the logic of the programme and take a critical perspective on what it is trying to achieve and how it does it.
6. The importance of evaluation questions in an evaluation design cannot be overstated. The temptation otherwise is to gather large quantities of data and produce sometimes technically sophisticated indicators which make little contribution to practice or policy. There is, of course, a problem formulating the evaluation questions in a way that they are likely to be able to be answered. While this is a technical question and this chapter has offered suggestions about how to formulate questions appropriately, there is here also the overarching concern for use. Ask questions that someone will find useful. However, use should not itself be defined too narrowly. We are talking here not just about the instrumental use of evaluation by managers. We are also talking of uses that citizens and civil society groups may make of evaluation in support of democratic processes and accountability.
7. We have specified in some detail the content and form of ideal Terms of Reference for an evaluation. This is part of the general question of design and the choices that can be made at the design stage which can influence the quality and direction of an entire evaluation. It should

be recognised that defining scope, clarifying the users of the evaluation and deciding the skills required for an evaluation team, are among the most important decisions made during the course of an evaluation.

8. It used to be common to regard the use of evaluation as being confined to acting on recommendations and final reports. It is now understood that evaluation use can be supported and occurs throughout an evaluation. Process use should involve stakeholders in evaluation thinking from the beginning. There are evaluations where the conclusions and recommendations are rejected but stakeholders, especially those involved in the steering committee find the evaluation useful. It can help them to clarify their own thinking and understanding and spark off innovative ideas for improvements. Promoting dialogue during the course of an evaluation is likely to ensure that when stakeholders receive reports they will be better prepared and receptive.
9. It is often easier for programme managers and those who commission an evaluation to confine judgements of evaluation quality to the outputs in reports of the evaluation itself. However, this quality control process provides few opportunities for learning and improvement in the way the evaluation itself is managed. A quality assurance perspective provides a context in which to explain the strengths and weaknesses of evaluations. It offers an opportunity for those who commission evaluations to learn how to improve in future.
10. Consideration should be given at an early stage to how evaluation findings will be used. Some use will stem directly from the findings and recommendations of the work. Evaluations can also be helpfully combined with other evidence to inform debates. The process of evaluation can bring benefits in terms of structuring inquiry and institutional reflection. Close attention to the factors that influence the use of evaluation work will maximise its contribution.

Chapter 3: Developing capacity for socio-economic evaluations

Evaluation capacity is part of institutional capacity

We have already painted an ambitious picture of what is expected of evaluation, how it should be organised and what it can deliver. In the previous Chapter, for example, we provided guidance to administrations that commission or use evaluations as to how they might design and implement an evaluation. This assumed that requisite institutional and administrative capacities were available. For our purposes institutional capacity is defined broadly to include legal rules (e.g., regulating employment and procurement) normative assumptions (e.g., about equal opportunities or open competition), governance (e.g., democratic accountability and divisions of responsibility between tiers of government and civil society) as well as administrative and organisational arrangements (e.g., how ministries are structured and resourced). In this Chapter we are concerned with how to create such capacity so as to make practicable the ambitions for evaluation that can contribute to socio-economic development.

Capacity cannot be created overnight nor is it created without costs. However, the potential benefits of evaluation are large enough to justify the investment and the recurrent costs needed to continuously innovate both in evaluation processes and products. It takes time to develop capacity and the necessary systems need longer-term nurturing to deliver sustainable benefits

Developing evaluation capacity has to be a shared concern of the wider policy community. Those responsible for policies and programmes must first be convinced of the need for evaluation. At the same time, even though their support is essential, they must not be allowed to capture the process. To begin with developing evaluation capacity includes those who manage and commission evaluations, those who have an interest in evaluation results at a policy and programme level as well as those who undertake evaluations. However, evaluation requires an institutional framework and has to be embedded in institutional arrangements. The capacity of public institutions to conduct evaluations is part of the wider requirements that the State must meet to address contemporary economic and social demands. Where evaluation capacity has been most developed is often in the very sectors that have conceived of it as an integral part of a much wider programme of public sector innovation and modernisation.

The need to build institutional and administrative capacity is a sometimes implicit but increasingly explicit transversal goal of socio-economic development policy. This goal in turn stems from two imperatives:

1. To overcome the inefficiencies of traditional public administrations by shifting towards a public management approach that draws on best management practice not only from the public sector but also taking on board lessons from the most successful private and non-profit organisations. Many of the assumed limitations to efficiency of public sector bodies are encapsulated in the so called New Public Management (NPM) movement which advocates reform, modernisation, decentralisation and privatisation among other solutions in the public sector. NPM aims to improve results, delivery and value for money.
2. To overcome the perceived distance of public bodies from society as a whole, and therefore open up policy making to the contribution of external stakeholders, civil society representatives and citizens (the global drive towards democratic governance).

The diffusion of evaluation can contribute to both of these imperatives.

What is evaluation capacity?

Evaluation capacity is multi-faceted and needs to be located at many different levels that reinforce each other:

- An Individual level consisting of necessary skills and competencies;
- An Organizational level of management arrangements and structures;
- An Inter-organizational level that bridges public and private bodies through networks, procedures and partnerships; and
- A Societal level that embeds evaluative thinking in civil society including professional organizations - as well as in the public sector.

Cutting across all these levels are institutional factors, regulations, laws, resources, norms, etc., that necessarily underpin organizational arrangements and mechanisms.

In an ideal situation one would expect that many of the following arrangements would be found:

At an individual level:

- There are people throughout government who have experience and skill in evaluation and this is reinforced and renewed by appropriate recruitment, training and professional development.
- Training courses and diplomas are available, delivered by universities, private training providers and professional bodies, open to evaluation practitioners and commissioners.

At an organisational level:

- Evaluation is routinely undertaken at each stage of policy and programming: from planning through to implementation and follow-up.
- Evaluation findings are integrated into decision-making when deciding what policy options to choose, how best to implement and when assessing what has been effective.
- Managers look to evaluation as one important input that will help them improve performance and manage for results.
- A regular flow of evaluations are commissioned that cover the broad spectrum of policies and programmes.
- There are follow-up procedures to ensure that evaluation recommendations are taken seriously and, where feasible, acted upon.
- There are procedures to accumulate evaluation findings and lessons learned so that programme managers and policy makers have an accessible evidence base and an organisational memory.

At an inter-organisational level:

- There is coordination through a network of dedicated evaluation units or functions - to ensure sufficient consistency in the way evaluations are commissioned, managed and executed across government and ultimately across the public sector.
- There are requirements that evaluations take place embodied in legislation, articulated policies and regulatory activity (e.g., audit or parliamentary reviews).
- There is a well-defined market with clear rules so that potential evaluation providers can organise themselves to respond to tenders, complete evaluation assignments on time, develop expertise and understand the priorities of commissioners of evaluation.
- There is a culture of evaluation that values professional standards, independence, learning from experience and evidence based policy.

At a societal level:

- Open and systematic dialogue is maintained between policy makers and evaluation specialists so that priorities for evaluation can be identified and scrutinised.
- There is an evaluation community of practice whose members may work for universities or consultancies or be independent evaluation practitioners but still share a common ethos and standards.
- Evaluation associations exist that bring together those who commission, provide and use evaluations and reinforce a culture of evaluation, disseminate good practice and safeguard the independence of evaluation functions and practitioners.
- There is an awareness of evaluation activity and outputs and a dissemination of reports and findings such that evaluations will be routinely used by various stakeholders (in parliaments, civil society, etc.) to support democratic accountability and transparency.

Developing evaluation capacity

In some countries evaluation capacity evolves organically over a long period of time. This was the case in the UK, Netherlands and Sweden. In most countries, however, there is a need for a strategy if evaluation capacity is to become useful within a defined time-scale. This was the case in most Member States joining the EU in 2004 and 2007. There are many key decisions to be made when starting to develop evaluation capacity in a strategic way. Among the most important are:

- Architecture: locating and structuring evaluation functions and their coordination;
- Strengthening evaluation demand: ensuring that there is an effective and well managed demand for evaluations;
- Strengthening evaluation supply: ensuring that the skills and competencies are in place with appropriate organisational support;
- Institutionalising evaluations: building in evaluation to policy making systems and across the broader policy system.

The "Architecture" of evaluation systems

It is important that those responsible for developing evaluation capacity are aware that they are developing a system in which the various elements will have to interact and reinforce each other. These elements will include, for example, responsible units or divisions in and across departments, procedures, human resource policies, information technology systems, etc.. In these circumstances a strategy needs to be somebody's responsibility. Usually ownership at least in the early stages of developing evaluation capacity is located in a dedicated unit within central government. This might be in a planning department, a national Treasury, a coordinating department, the prime minister's office or a department responsible for EU Cohesion Policy.

Two main architectures have been most common at the start-up stage of developing evaluation capacity:

- A central evaluation unit or function which from the beginning is also associated with a network of other ministry based or sectoral or programme based units. The central unit provides an umbrella organisation within which basic procedures, guidelines, commissioning and public procurement protocols are agreed. Once a system is established the central unit continues to orchestrate but other units increasingly have an equal voice and represent their own priorities within a network.

- A central unit that is located in one department at the outset. It is intended once such a unit is established to diffuse the model that has been piloted and proven itself. Such diffusion might be formalised as when it is decided centrally that all ministries will in future have evaluation units. Alternatively a more informal process might be chosen: diffusing good practice and encouraging replication.

There are advantages and risks associated with both approaches. A central evaluation unit or function has to strike a fine balance between ensuring that a coherent system evolves - whilst not centralising too much. A balance has to be maintained between adapting systems to local circumstances and ensuring consistency. Centralisation in one coordinating ministry can risk undermining ownership and support for evaluation elsewhere in government, the public sector or among other stakeholders. Too much de-centralisation on the other hand - for example having different coordination arrangements for different programmes, policies or territories, risks incoherence.

Once an evaluation system is established there are further options for location and structure. For example, some ministries or public agencies may concentrate their evaluative efforts in their own central unit, whilst others may rely on internal networks with one or two individuals in a central node and others dispersed across different policy departments, units or divisions.

If evaluation is to become a valued function within the public sector it must itself mirror the architecture of governance within the country concerned. For example, if regional or federal tiers of government are important, then evaluation must also be embedded in these sub-national governance structures. Although this may not be where evaluation is first located, new evaluation systems need to be designed from the beginning with a vision of how they will be adapted more generally.

Strengthening Evaluation Demand

1) Governments as a source of demand for evaluation

The demand for evaluation, the first push that kicks-off the process of capacity development, can come from internal or external sources. In many European countries Cohesion Policy regulations provided the first impetus so the demand was at first external. This was true for southern European countries such as Greece, Italy and Portugal and most of the EU12. This external demand is in part an outsourcing of the responsibilities of European institutions that are obliged to be accountable. The European Commission has obligations to the Member States, the European Council, European Parliament and the European Court of Auditors to ensure that the monies made available for socioeconomic development are spent wisely and that the policy goals of better economic, social and territorial cohesion are achieved. Because the same governments that provide the financial resources are the recipients of the money, it is only natural that the evaluation is entrusted at least partly to them instead of only being carried out at a European level. Similar external demands for evaluation are common in international development. Donors, whether individual countries or multilateral agencies such as the EU itself, international banks (World Bank, European Bank for Reconstruction and Development, etc.) or even UN agencies, all require that beneficiaries of funds demonstrate that they have used monies wisely.

Alongside the accountability motive, external actors who initiate a demand for evaluation are also interested in effectiveness. However this motivation is even stronger when the impetus for evaluation is internal.

Many of those who have embarked on evaluation capacity development for domestic reasons have done so as part of a commitment to an effective, efficient, transparent and accountable public sector. Evaluation in Canada, the Netherlands, Sweden and the UK did not begin with external demands. Rather evaluation demand was driven by successive waves of internal reforms of the public sector

aimed to improve its management and effectiveness and new notions of governance. This was also true for early stages of public sector reform in Australia, which was mainly driven by a results-based philosophy also reflected in the emphases given to evaluation and its utilisation.

Even when the driving force behind the establishment of evaluation comes from within national borders, a certain degree of external scrutiny is likely. This may take several forms:

- Parliaments, at national, regional or local level, which seek to make government responsible for the efficient and effective implementation of the decisions they have taken;
- Courts of Auditors and similar bodies, wishing to expand their role from the verification of legality and financial control to performance audit, including notions of effectiveness, efficiency and value for money;
- Central government that finances the activities or the investment of sub-national authorities and make it a condition that the latter introduce evaluation in their practices and open themselves up to some form of external scrutiny.

If the impetus for evaluation is entirely external (or even externalised by government departments or agencies to national scrutiny bodies) the foundation for building evaluation capacity is likely to be weak. There has to be a recognition that evaluation is more than an obligation to someone else for demand to be taken seriously. This principle is central to debates about the need for regulatory or legal force behind evaluation. There are those who argue that without a clear legal demand backed up by appropriate compliance mechanisms evaluation will not be taken seriously. On the other hand there is also a strong argument that regulation without cultural, managerial and administrative change will not be effective. There will be token compliance rather than wholehearted adoption of new practices.

There are a number of reasons why systems based exclusively on rules and regulations may not work:

- They depend for their implementation on some kind of sanction. However the ability or willingness of the relevant political authorities whether European Commission, Member States' central government or sub-national authorities, to exercise such sanction is often limited. These limitations may derive as much from political considerations as from legal or administrative capacity.
- With limited likelihood of sanction, evaluation requirements are taken less seriously and evaluation processes and documents tend to be disregarded by the same bodies that prescribed them.
- Managers pre-occupied with their day-to-day responsibilities will minimize their time commitment to evaluation if it is not seen as directly relevant and useful to them. It is unlikely that evaluations designed to meet the needs of external parties will be relevant and useful.

The implications of this latter point are profound. Strengthening demand for evaluation requires the mobilization of a broad coalition of internal stakeholders including managers and policy makers. It cannot simply be created by one or two committed evaluation champions or by deciding to establish a small central unit. Once such a coalition has been established then further procedures and policies can be developed that will reinforce initial demand. Many of the activities identified in Chapter 2, concerned with the commissioning and management of evaluation become relevant. However the strongest boost will come from embedding evaluation in the routine practices of public agencies or government ministries.

2) Civil society as a source of demand for evaluation

A sustainable evaluation system cannot be confined to governments or even the public sector. Civil society can also demand evaluation. Strengthening evaluation is often thought of as a top-down national

or even supra-national process. Accountability and learning take place among policy makers and other actors at lower levels are expected to comply with the requirements of the policy community. In terms of contemporary policy thinking this top-down view is anachronistic. Participation, active involvement, responsibility and transparency all assume that civil society actors are engaged with public policy making and delivery. Capacity development has to include civil society actors if it is to be true to the logic and values of public sector reform. This can be achieved through public consultations, seeking inputs into priorities for evaluation plans and, as suggested previously, ensuring a broad representation of interests on steering committees and advisory groups.

Supply side

1) Starting Conditions and Traditions

However strong the demand for evaluation, the response to that demand will depend on pre-existing resources, skills and institutions. To a great extent it will depend on the skills, resources and institutions outside the administrative and public management system. There is not a single way to improve evaluation supply. Three approaches are common:

- To build up relationships with educational institutions, in particular, universities;
- To develop and strengthen an independent community of consultants; and ▪
To support the development of a professional evaluation community.

2) Links with the education sector

Educational institutions and universities in particular are important for the supply side of evaluation for two reasons. First, universities are in many countries the natural place to locate training and education programmes necessary to develop evaluation skills and knowledge. Second many working in educational institutes already have much of the knowledge of social and economic research, data gathering and analysis and report writing that are needed by evaluators. Educational and research perspectives need to be supplemented with other methods, theories and practices to become useful in evaluation, however.

Establishing a working relationship with universities and other educational institutions can serve to stimulate the offer of formal training programmes, continuing education and the creation of centres of excellence linked to research and able to disseminate knowledge. Courses can be directed at graduate students in social sciences and economics, specialists in policy and regional studies and practitioners in socio-economic development. Depending on the target group, it may be best to create new specialised courses or to internalise evaluation modules, e.g., theories and methods of socio-economic evaluation into the formal curricula of economists, engineers, sociologists, planners, etc..

It is important when planning education and training programmes to recognise practitioner aspects of evaluation. Working in a client and contract environment, assessing what will be most useful for programme managers, planning and team management and dealing with multiple or even conflicting interests tend to be more common in an evaluation rather than in an academic environment. Practitioner skills are needed as well as academic knowledge. This can be achieved by including within curricula workshops, training seminars, guest lectures from practitioners, work placements with existing evaluation consultants, study visits to administrative bodies, etc..

In some countries there is a tradition of universities and institutes actively engaging in policy research. In these countries there will be a pool of skills that can be tapped into to increase the supply of evaluators

in the relatively short-term. However, even in such countries, this can require cultural change among those who are used to a more academic pace and style of research. In these cases the increase in evaluation supply will take time to mobilize, beginning perhaps with awareness-raising seminars and then specific small scale assignments that will allow potential evaluators to test out new ways of working in a policy environment.

3) Independent Consultants: encouraging new suppliers

In many fields of evaluation there is an international market for evaluation. Large international consultancies are used to bidding for work in a number of countries, not just where their headquarters might be located. It is therefore common for early entrants to new evaluation markets to come from among these international consultancies. They have the potential to import a wealth of international and policy specific experience as well as experience of working with policy makers and programme managers. However, it is important that there should be a home grown evaluation consultancy sector, locally based consultants who are familiar with national institutions, policies and languages. Of course international consultants will often also recruit in national counterparts, thus encouraging international exchange even within a single evaluation team. Such joint working can be an effective method of skill transfer as well as knowledge dissemination. This is widely recognised in different evaluation sectors. For example in international development programmes it is often a requirement that evaluation teams include experts from the northern and southern hemispheres.

There are various approaches to encouraging the emergence of national consultancy suppliers. These include:

- Commissioners of evaluation insisting on consortia or partnership bids that always include some local consultants;
- Scaling evaluation contracts in ways that relatively small low-risk evaluations can be undertaken by national new entrants to the evaluation market;
- Ensuring that procurement and financial requirements associated with bidding for evaluations are not too restrictive (responses required within short time periods, short periods also for bank guarantees and years of audited accounts);
- Emphasising in ToR technical and know-how criteria rather than complex administrative procedures with which less experienced consultants may not be familiar;
- Holding briefing meetings with potential consultants to answer questions and encourage bids in a competitive environment; and
- Support for networking among relatively isolated evaluation consultants so as to encourage team-building, consortia formation and other professional networks.

4) A professional evaluation community

It is in the interest of those who commission and undertake evaluations that they conduct themselves within a framework that encourages high standards. This is what is understood when evaluation is described as a profession. Professional standards are essential for evaluators to be regarded as credible and for evaluators to be granted necessary independence by those who commission and use evaluations. This is especially important in the face of the different and sometimes very powerful interests involved in programme and policy evaluation. Many evaluators have experienced attempts to interfere with how evaluations are conducted. This can take various forms:

- Unbalanced ToR;

- Undue influence on an evaluation process by those who are responsible for the programme being evaluated;
- Attempts to pre-select an evaluation team to include those who will come up with the right answer;
- Pressure to reshape findings and conclusions to be less critical and more acceptable to the main stakeholders; and
- The provision of limited or selected data.

In reality it is in no one's interests for this to happen. For a modern administration the benefits of impartial and trustworthy evaluations far outweigh the apparent benefits of unchallenging and safe reports. However, it is only with the development of an evaluation culture that this becomes apparent.

Professional standards in evaluation are usually assumed to include a commitment among both evaluators and commissioners to:

- Using the best/most appropriate methods;
- Self development and upgrading of skills;
- Taking seriously the interests of stakeholders and policy customers, including programme managers;
- An independent stance that is impartial and true to the evidence collected; and
- Ethical behaviour towards various parties: commissioners, those who provide information and data, potential beneficiaries of programmes and policies.

There are various sources of professional strength and independence. They include:

- An evaluation culture that reinforces professional norms of behaviour;
- Respect for independence by those who commission evaluations;
- Ethical codes on which there is a consensus and which are therefore widely recognised and disseminated;
- Independent institutions within which evaluators work which can lend their judgements greater weight and allow them to resist external pressures;
- High quality education and training and professional development;
- Professional associations and societies that bring together all those involved in the evaluation community where experience can be shared and practical problems discussed; and
- The development and application of quality standards for practitioners and commissioners of evaluation - such as those suggested in the previous Chapter.

In practice many of these desirable attributes of a professional evaluation community can be encouraged under the umbrella of a national evaluation society or associations. These typically bring together people from different institutions (academia, public administration, research centres and consultancies) with different disciplinary expertise (sociologists, economists, planners, political scientists, etc.) and different fields of specialisation (social services, education, research and development and of course socio-economic development). Such societies have become widespread across Europe in recent years. Such societies provide an opportunity for cross fertilisation of ideas, the discussion of shared problems outside of the setting of particular evaluation contracts, the dissemination of good practice and the emergence of common standards and codes of ethics.

List of Evaluation Societies and Associations:

- European Evaluation Society: <http://www.europeanevaluation.org/> ▪ Danish Evaluation Society: <http://www.danskevalueringsselskab.dk/>
- Finnish Evaluation Society: [http://www.eukn.org/E_library/Economy_Knowledge_Employment/Training/Training/Finnish Evaluation Society FES](http://www.eukn.org/E_library/Economy_Knowledge_Employment/Training/Training/Finnish_Evaluation_Society_FES)
- French Evaluation Society: <http://www.sfe-asso.fr/>
- German Evaluation Society: <http://www.degeval.de/>
- Italian Evaluation Society: <http://www.valutazioneitaliana.it/>
- Spanish Evaluation Society: <http://www.sociedadevaluacion.org/website/> ▪ UK Evaluation Society: <http://www.evaluation.org.uk/>

Institutionalising evaluation

Much of this Chapter has been concerned with the initiation of evaluation capacity and with arguments about why this is important both for socio-economic development and for good public management. Institutionalisation is more concerned to embed and deepen evaluation within public sector institutions. This usually involves:

- extending evaluation more widely within the public sector, and
- greater integration of evaluation processes and utilization into policy making and programme management.

It is likely that the introduction of evaluation will first be concentrated in one part of a government. Extending evaluation spreads evaluation practices to other ministries or programmes and to other institutions. In mature evaluation systems evaluations are used by national audit bodies and parliaments as well as by ministries, policy makers and programme managers.

Integrating evaluation more closely into policy making and programme management is likely to involve:

- The joint planning of policy and evaluation through multi-annual cycles of consultation resulting in multi-annual or departmental plans that include evaluation;
- Systematic follow-up of evaluation findings and recommendations such that it becomes difficult for middle managers to ignore evaluations as they need at least to say why they are not following recommendations;
- The creation of forums where policy makers and managers and evaluation managers meet to discuss evaluation findings and lessons learned; and
- The extension of evaluation throughout the policy cycle from option identification to planning, programming, implementation and reprogramming to the design of the next phase of policy.

Creating an evaluation culture suggests that a well developed evaluation system is more than just a structural arrangement. For example:

- There is a commitment to learning lessons and improvement;
- There is avoidance of a blame-culture which discourages learning;

- Policy makers are committed to evidence based policies in the broadest sense;
- There is a commitment to excellence, high standards and continuous improvement;
- Evaluation is used as one element in a general move towards transparency and multiple accountabilities to citizens and communities as well as to Ministers and parliaments; and
- The government and public sector is committed to continuous adaptation to becoming a learning organisation.

Canada is widely seen as a country which has been successful in institutionalising evaluation into its public management and policy making systems at federal (i.e., national) level. Overall responsibility for evaluation coordination is vested in the Treasury Board for Canada which must provide central direction for evaluation in the Government of Canada and, to that end, should¹:

1. establish a Centre of Excellence to provide leadership, guidance and support to the practice of evaluation;
2. use evaluation results where appropriate in decision-making at the centre;
3. set standards; and
4. monitor evaluation capacity in the government.

Distinctive features of Canadian evaluation policy are that it:

- Relies heavily on the devolution of responsibilities to develop and use evaluations to government departments (ministries) and to managers within these departments,
- Focuses on results based management as a component in public management this, not as a backward looking accountability focus, although accountability is also emphasised

Assessing Evaluation Capacity Building:

The authors of an international comparison of evaluation capacity used the following nine criteria to assess the extent of evaluation capacity:

- 1. Evaluation takes place in many policy domains**
- 2. There is a supply of domestic evaluators in different disciplines**
- 3. There is a national discourse about evaluation**
- 4. There are professional organisations of evaluators**
- 5. There is a degree of institutionalisation of evaluation in government**
- 6. There is a degree of institutionalisation in Parliament**
- 7. There are many institutions of evaluators conducting evaluations within each policy domain**
- 8. There is an evaluation function within the national Audit Institution**
- 9. There is a balance between evaluations that assess outcomes and those that assess outputs and processes**

International Atlas of Evaluation, Editors Furubo, Rist and Sandahl, Transaction Publishers, 2002

Institutionalisation of evaluation is a continuous process which over time has to be integrated with other planning and assessment tools (e.g., impact assessments, project planning techniques) and other

¹ <http://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=12309>

channels for identifying and disseminating and implementing good practice (e.g., public consultation, decentralisation to local stakeholders). At the same time potential barriers to institutionalising evaluation cannot be ignored:

- Evaluation is stronger when it is seen as an integral part of institutional development and public sector reform. If these broader organisational and cultural changes are not pursued in parallel it will be more difficult to institutionalise evaluation.
- A coherent institutionalisation process requires both financial and human resources. Institutionalisation can be undermined if there is insufficient investment in the skills and management systems for specialist evaluation units.
- High-level political commitment is also important in the evaluation institutionalisation process. This is especially so when such a process inevitably takes time and needs to be built up in stages. Changes in direction when officials or ministers change can be a barrier to the successful building of evaluation capacity

As in other Chapters, key lessons are summarised below in terms of golden rules for those seeking to strengthen evaluation capacity.

Golden Rules

1. Evaluation is most useful and used when it is embedded as a function in supportive administrative and institutional systems that seek also to include civil society stakeholders. This is what is meant by evaluation capacity.
2. Evaluation capacity can develop organically over time but in most national and international settings it needs to be planned for, consciously introduced and appropriately resourced.
3. The principle of independence and impartiality is an essential principle to build in to evaluation functions given the vested interests that are likely to seek to influence the conduct of evaluations and their conclusions.
4. Evaluation capability consists of elements at individual, organisational, inter-organisational and societal levels that should reinforce each other. Evaluation capacity should be designed as a system.
5. Evaluative activity needs to be coordinated across administrations to ensure coherence and provide a basis for skills development and professionalisation. At the same time there must be sufficient flexibility in the procedures and approach to adapt to the specificities of policies, programmes and territories.
6. Evaluation capacity may begin by focusing on particular programmes, e.g., those funded by the EU or those taking place at a particular tier of government (e.g., central government departments) but should be designed so that evaluative activity can be extended for example to include nationally funded programmes or regional as well as central government activity.
7. Evaluation should be configured to meet real decision-making and management needs that will improve the effectiveness and efficiency of public action and yield concrete benefits for citizens and stakeholders. Whilst evaluation will inevitably be useful to fulfil external accountability needs, it should never be primarily regarded as an external obligation.
8. The inclusion of civil society and the encouragement of participation in evaluation, whether through consultation or direct representation will encourage different interests (e.g., community representatives, professional groups and the private sector) to demand and use evaluation, thus strengthening accountability, increasing transparency and reinforcing social capital.

9. Evaluation capacity must include attention to the supply side, promoting a skilled and professional community of practitioners and the institutions, networks and markets that will ensure that evaluation assignments are bid for and delivered to a high standard.

Among the key underpinnings of a professional evaluation community are education and training for evaluators and quality standards and ethical codes agreed and shared among the commissioners and practitioners of evaluation including academics and consultants. These supply-side qualities are often best developed through national evaluation societies with open memberships and a commitment to high standards.