Three layers of quality assurance: would this help provide EU policy makers with the evaluative information they need?

Jacques Toulemonde, Hilkka Summa and Neil Usher

In a further version, this paper will be part of a book edited by Schwartz, R. and Mayne, J. at Transaction Publishers under the (provisional) title of ‘Assessing Evaluative Information’
Abstract

This chapter describes three different forms of quality assurance that have applied to the policies and programs of the European Union, either at large scale or in an embryonic manner. It comments on their relations to each other and sketches how a comprehensive 3-layer system could be developed from the current practices.

The first layer applies to external evaluators’ reports. It is handled by the evaluation managers (or evaluation function) in line departments (DGs). This assessment relies upon an quality assurance system which is internal to the department. The system is both compatible with internationally recognised standards and finely tuned to the specific needs of the policy area. At this level, evaluative information is assessed in both a formative and summative way, each evaluation report being assessed mainly in terms of substance.

The second layer of quality assessment applies to the evaluation system which is run in each department. Therefore it looks like systems audit but it may range from a formal audit to an informal peer review. It may be operated in a centralized or decentralized way or both. Some internationally agreed standard also apply at this level. Quality assessment at this level may be either formative or summative. In the later case, it is run as a formal audit and used to strengthen the credibility of the first level of quality assessment, something which will be increasingly needed as long as evaluations become subject to public debates.

To a large extent the two first layers consider quality with the standpoint of the producers of evaluative information, since they apply standards resulting from a consensus among experts.
On the contrary, the third layer applies to the whole set of evaluative information received by decision-makers, with their standpoint, that is to say with the consumer’s standpoint. Quality standards at this level should mainly derive from users’ needs and not from the evaluation profession. At this third level, external audits have to question the value added of evaluation in order to ensure that the two previous layers of quality assurance remain user-oriented.
Contents

European policies and programmes are periodically submitted to external evaluation......... 1

Assuring the quality of external evaluators’ reports ................................................................. 3

Quality criteria: the MEANS grid .......................................................................................... 3

Assuring the quality of evaluation reports: the case of Agriculture Policy......................... 9

Assuring the quality of evaluation management ...................................................................... 19

Setting up criteria: the Good Practice Guidelines ................................................................. 19

Auditing evaluation systems ................................................................................................ 27

Assessing the quality of information provided to policy makers............................................. 32

Evaluation and the European Court of Auditors ................................................................ 32

Towards an integrated system of quality assurance ............................................................... 38

References ............................................................................................................................ 41
This chapter describes three different forms of quality assurance that have applied to the policies and programs of the European Union, either at large scale or in an embryonic manner. It comments on their relations to each other and sketches how a comprehensive 3-layer system could be developed from the current practices.

**European policies and programmes are periodically submitted to external evaluation**

The practice of evaluation in the European Union dates back to the late 1970s with first steps in the fields of Development Aid and Research and Technology policy. Regional Development policy followed in the late 1980s. In 1977, a Financial Regulation stipulated that all actions occasioning expenditure should be submitted to an evaluation. However, explicit rules for regular evaluation of expenditure programmes were established in 1996 only. The administrative reform of the European Commission which started in 2000 extended the principle of periodic evaluation to all kinds of activities, including policies which not necessarily involve spending and which may be designed with less precise objectives and time frames than programmes. In the first years of the millennium, it can be said that all European policies and programmes are submitted to periodic evaluations (Summa & Toulemonde, 2002), although the extent to which evaluation practices are established and regular varies from one policy area to another.

European policy-making involves a complex sharing of powers between the Council, the Parliament and the Commission, the latter taking the lead when it comes to initiating, implementing and evaluating policies. Every year, about 100 evaluations are launched by the departments of the Commission in charge of the various policies (Directorates General,
hereafter DG). Most of these are contracted out through calls for tenders to external evaluators, either consultants or academics. Most of the reports are publicly available, increasingly on the Internet.

This evaluation system produces information that is mainly used by policy designers and managers within the Commission. Evaluation reports typically focus on the achievement of expected outcomes and contain recommendations for improving the design and management of current policies and programmes.

The Council and the Parliament have supported the development of evaluation from the very beginning, but their interest into evaluative information has increased recently. There is a growing expectation for evaluation to serve the purposes of accountability. Evaluation is now increasingly put into the context of governance, as compared to the situation where internal, managerial use was the main focus. This puts more pressure on assuring the quality of evaluative information.

Even if the evaluation system is already well established, quality assurance of evaluative information is still embryonic. Concerns for the quality of evaluative information have repeatedly been raised in Commission wide policy documents, by individual DGs and by the external auditors. As a result, some quality assurance instruments and procedures are already in place, although they do not cover all evaluations in a systematic way. What is described below is therefore not a stable and consistent system but evolving practices.

Nevertheless, the emerging quality assurance instruments and procedures seem to
complement each other and could be merged into a consistent scheme of a quality assurance. The authors have chosen to describe these emerging practices as a three-level quality assurance scheme which combines:

1. DGs’ efforts to secure the quality of individual reports produced by external evaluators,
2. Commission wide mechanisms assuring the quality of evaluation systems in the DGs,
3. External Auditors’ assessment of evaluative information from the standpoint of policy makers.

In the concluding section the three levels of quality assessment will be drawn together, with an aim to better understand how they interact, what added value each layer has and how they could be combined in order to improve evaluation and eventually policy outcomes.

**Assuring the quality of external evaluators’ reports**

*Quality criteria: the MEANS grid*

**A step by step construction**

In 1995, the Directorate General in charge of EU Regional Policy launched an international workshop with an aim to set up a quality assessment method. Six experts from several European countries attended successive meetings. They took stock of the main assessment frameworks that were available in Canada, France, Netherlands, UK and USA. They finally agreed upon a quality assessment grid that was subsequently disseminated within the Commission and in the Member States (European Commission, 1999a pp 181-193). This
grid is known as the MEANS grid, from the name of the program that supported the workshop.

The MEANS grid has been subsequently tested in a series of action-training seminars with dozens of European officials across the Commission. Several DGs started to use it partly or totally, as well as some national administrations (Uusikyla & Virtanen, 2000).

Among the various uses of the grid, it is worth describing that of the DG Agriculture, since it has gone very far in terms of systematic assessment of the quality of evaluation reports. A dozen reports were assessed from 1999 to 2001. The MEANS grid has proven to be an operational tool, although marginal improvements have been introduced to the original version. In its present state, the grid includes nine criteria and a rating scale for each of these. A high-quality evaluation report is described according to these nine criteria in the following way:

Criterion 1: Meeting needs

- The evaluation report adequately addresses the information needs of the commissioning body. It answers all questions included in the terms of reference in a way that reflects their stated level of priority. It satisfies incidental information needs that have arisen during the evaluation process.

Criterion 2: Relevant scope

- The report precisely describes what is under evaluation, for instance in terms of budget lines, time period, geographical areas.
• The report adequately covers the rationale of the policy, its history and its stakeholders. The set of inputs, implementation mechanisms, outputs, intended outcomes and other consequences are examined fully, including long term and unexpected consequences and policy interactions.

• The report pays sufficient attention to the rationale of the policy, to the clients’ views and to policy outcomes. It goes beyond management and implementation issues.

• The report addresses issues that are potentially relevant for policy reform or management improvements, e.g. comparing several policy instruments or judging efficiency.

Criterion 3: Defensible design

• The evaluation method is clearly described. Major options are explained and justified, including the risks associated with alternative options. The stated method has been actually applied.

• The evaluation method is appropriate and adequate to answering the evaluation questions.

• The report clearly states the strengths and weaknesses of the method and the subsequent limitations that may restrict the use of evaluative information.

Criterion 4: Reliable data

• This criterion does not assess the quality of pre-existing information but how the external evaluator has managed to find, to produce and to use information.

• Primary and secondary data are sufficiently reliable for their intended use.

• In the case of surveys, the techniques of data collection are clearly stated and the report demonstrates that they have been applied in accordance with their own quality standards.
(e.g. selection of interviewees, size of samples, test of questionnaires, rules of confidentiality, etc.).

- Secondary data are cross-checked whenever it is relevant and their sources are quoted.
- All potential biases are identified, analyzed and corrected as far as possible by recognized techniques. The remaining risks to validity are stated.
- The report is free of useless data (i.e. data that do not substantiate the conclusions).

**Criterion 5: Sound analysis**

- Data are appropriately and systematically analysed or interpreted according to the state of the art. Analyses and interpretations are sufficiently reliable for their intended use.
- Major cause-and-effects relationships and underlying assumptions are made explicit. Critical exogenous factors have been identified and taken into account.
- In case of comparison (e.g. before-after, with-without, etc.) or statistical analyzes (e.g. regression, factor analysis etc.) techniques are clearly stated and the report demonstrates that they have been applied in accordance with their own quality standards (e.g. statistical tests, validity thresholds, attrition biases, etc.).
- The report is free of useless analyzes (i.e. analyzes that do not substantiate the conclusions).

**Criterion 6: Robust findings**

- The report provides stakeholders with a substantial amount of new knowledge (findings).
- Findings are clearly identified. They follow logically from, and are they are justified by, data, interpretations and analyses through logical reasoning that are carefully described
and do not contradict each other. Logical reasoning are developed as far as possible and necessary.

- When relevant, the report indicates which findings are generalisable and under which conditions. Generalisable findings are presented in order to be easily acknowledged and used by the readers.

Criterion 7: Impartial conclusions

- This criterion does not assess the substance of the conclusions but how the external evaluator has built up and justified the conclusions.

- The report includes clear and accurate conclusions. These conclusions go further than findings in the sense that they include value judgements.

- Judgements are fair, unbiased by personnel or stakeholders’ views. They impartially take into account all legitimate standpoints. Conflicting issues are presented in a balanced way.

- The conclusions are based on explicit and agreed judgement criteria and benchmarks. There are no discrepancies between stated criteria and benchmarks and those which have been actually applied.

- Key conclusions are highlighted.

Criterion 8: Useful recommendations

- Recommendations are not mixed with conclusions, but they are derived from them.

- These recommendations are detailed enough and operational. The report indicates that practical constraints have been identified and taken into account (e.g. regulations, institutions, budget).
• Key recommendations are highlighted. If relevant, they are presented in the form of alternative scenarios.

Criterion 9: Clear report

• The report is logically structured. It includes a short summary and appendices. The size and style of the different parts make them accessible to the various categories of readers.

• The summary reflects the key findings, conclusions and recommendations in a balanced and impartial way.

• Only appendices contain technically difficult information that is not accessible to all policy makers and citizens.

Rating and combining criteria

In the original MEANS grid, all nine criteria are rated separately on four-mark scales, two marks being positive and two negative. On the basis of the first experiences of using the grid DG Agriculture added an intermediary neutral score. The following scale is now used:

(4) Excellent: with respect to criterion X, the report deserves to be quoted as “best in class”.

(3) Good: with respect to criterion X and considering the constraints and context of that specific evaluation, the quality of the report is good.

(2) Satisfactory: with respect to criterion X and considering the constraints and context of that specific evaluation, the quality of the report can be accepted.

(1) Poor: with respect to criterion X, the report should have been significantly better, but
this weakness does not justify to reject the whole work.

(0) Unacceptable: with respect to criterion X, the weakness of the report makes a major problem which cast doubts about the quality of the whole evaluation.

The quality assessment grid does not include any weighting of the nine criteria. The overall quality is to be rated by freely considering the nine individual ratings, in the light of relevant contextual factors like the constraints, difficulties or complexities that the external evaluator had to face.

Assuring the quality of evaluation reports: the case of Agriculture Policy

A well established process of quality assurance

The Agriculture DG is responsible for policies that account for the largest individual part of the budget of the European Union. An Evaluation Unit was set up in 1998 and charged with planning and managing evaluations of measures covered by the Common Agricultural Policy. So far all of these have been contracted out to external experts through open calls for tenders.

The Evaluation Unit manages a given evaluation with the help of a specifically convened group called an “evaluation steering group”. A steering group typically gathers six to eight times for the purpose of drafting the terms of reference, selecting an external evaluator, interacting with this evaluator during the course of the study, deciding whether or not to accept the final report and endorsing some or all of the evaluator’s recommendations. A typical steering group gathers European officials from the DG Agriculture Evaluation Unit, from the Operational Unit which manages the measures under scrutiny, and from other units
and DGs concerned with the area in question.

Since 1998, the Evaluation Unit has paid constant attention to the quality of external evaluators’ reports. At first, only the final reports were subject to a summative quality assessment with the help of the MEANS grid. However, it soon became clear that formative quality assurance throughout the evaluation process was to be preferred to avoid a judgement of poor quality at the end of the process.

From 1998 to 2001, twelve external evaluations have been subject to a step-by-step quality assurance process that has taken the following form. First, the quality assurance system is mentioned in the terms of reference of each evaluation and a detailed description of the rules of the game is publicly available on the web (2002: www.europa.eu.int/comm/agriculture/eval/guide). Second, the selected evaluator is briefed about the quality criteria and the quality assurance process. Third, the successive interim reports are subjected to quality assessments by the steering group.

Although an interim report does not lend itself to a complete assessment, some of the quality criteria may apply at least tentatively. During the inception phase of the works, it is possible to assess the proper understanding of the information needs (criterion 1), the proper definition of the evaluation scope (criterion 2) and the appropriateness of the method (criterion 3). Reliability of data (criterion 4) can be assessed quite early in the process, at least for the first waves of data collection. At mid-term, external evaluators are required to provide their draft findings and a provisional conclusion for at least one of the evaluative questions. This allows for a partial assessment of criteria 5 (sound analysis), 6 (robust findings) and 7 (impartial
At each stage, the Evaluation Unit and the steering group may require improvements or even reject the interim report for quality reasons, which did happen once.

A comprehensive, formal quality assessment is usually carried out on the draft final report. All criteria of the assessment grid are rated and the ratings are complemented with detailed explanations. A specific assessment may be made for a key evaluative question in addition to assessing the report as a whole. Cases of extreme ratings like “unacceptable” or “excellent” call for stronger justification with examples taken from the report.

A draft for the quality assessment is prepared by the responsible manager in the Evaluation unit, and then discussed by the steering group. Once the steering group has reached consensus on the assessment, it is presented to the evaluator. Practice has shown that significant quality improvements are still possible at this late stage of the evaluation process. In this case the steering group will revise its earlier quality assessment, check the ratings of the criteria and prepare final summative comments. The external evaluator is given a right to comment on that assessment. Both the final quality assessment and the evaluator’s possible comments are published together with the report. Most evaluation reports are made available on the internet together with the quality assessment.
Box 1: An example of quality assessment: Evaluation of the Common Organisation of the Markets in the sugar sector

In 1999, the Directorate General for Agriculture launched an evaluation of the measures in the sugar sector. The objective of the study was to examine the impacts of this regime, in particular impacts on the EC and the world sugar market, on the income of sugar beet producers and on the distribution of sugar beet production and processing between the EU member states. The group in charge of steering the evaluation process comprised officials from the evaluation unit, the operational unit in charge of implementing the sugar regime and from some other EC Directorates General. The group drafted a number of evaluative questions, the first of which being “Did the regime allow for supplying sugar to EU industry and EU consumers at a reasonable price?” The evaluation was contracted out to an external evaluation team which delivered its report in September 2000.

Successive quality assessments were undertaken along the evaluation process and the final report was eventually assessed in a summative way. The following extracts illustrate how the criteria were used in the final assessment:

Criterion 2 – Relevant scope – Excellent: “The report treats the different impacts in a rather exhaustive way taking into account the context of the Common Agricultural Policy, analysing relations with third countries and differentiating the analysis by Member State or region in several cases..... Unexpected policy interactions and consequences of the CMO for sugar are mentioned...”

Criterion 5 – Sound analysis – Poor: “... there is a lack of explanation of methods used in particular when quantitative methods are applied... The main point of criticism is that there are important elements of analysis which are difficult to follow because the assumptions on which they are based are not laid down and explained sufficiently. These assumptions are mainly based on economic theory (... assumption of ideal and functioning markets) without empirical verification for the concrete case of the sugar market. In principle it is useful .... to refer to economic theory, but when this is done one must check whether reality corresponds to the hypotheses developed in theory before conclusions can be drawn. In the report however this verification is not done sufficiently or at least not explained clearly enough. Examples for this are laid down in the annex to this assessment.” The examples include analyses related to: (1) rise of instability on the world market due to the EU regime, (2) sugar production in the absence of the regime, (3) security of sugar supply which could be assured through imports from the world market.

This assessment has been published together with the report, both documents being accessible through the Europa web site. The evaluation team was given the right to comment, and its reply is also published on the same site. The evaluation team endorses the overall quality assessment but responds to the critical assessments with detailed additional reasoning, for example in the following way:
**Criterion 5 – Sound analysis:** “Indeed these conclusions could not be based on empirical verification, because the counterfactual situations do not exist in reality. However, those conclusions are .... [based] on logical reasoning and generally accepted knowledge about how markets normally function. .........(3) The statement that the security of sugar supply can be assured easily through imports from the world market is based on .... the production and export potential of the low cost cane sugar producing countries such as Brazil, Australia, Thailand, south Africa, etc. Indeed their production potential has not been documented in the report, because it is a widely accepted fact by sugar industry specialists and documented in various articles and studies. ... It is likely that these countries can develop, over a number of years, a production potential to meet the demand of the EU. At least availability of land and labour would not be a problem.”

*Report, quality assessment and evaluator's comments are available on the Europa website*

**Practicalities of quality assessment**

Preparing the formal quality assessments of the draft final and final reports is time-consuming and usually takes several working days to complete. First the official in charge of managing the evaluation prepares a draft and discusses it with colleagues in the evaluation unit. Sometimes separate ratings of the criteria are done by two or three colleagues, and individual assessments are discussed until consensus is reached. The evaluation manager then records the agreed comments. Finally, the draft ratings and comments are submitted to and discussed in the steering group and endorsed by it after any necessary changes. Assessing a report often requires several in-depth readings, which means that a considerable investment of time is needed from all those who contribute to the task.

In practice, individual assessments often converge easily. This may be explained by the robustness of the grid, but also by the fact that members of the evaluation unit share a common culture. Disagreements of the quality assessment in the steering groups have also been rare. During the process of successive assessments of interim reports a common view of
the strong and weak points usually develops.

The assessors have faced some recurring difficulties in using the grid. For instance when the evaluation report relies on a qualitative design, they feel less confident in their ratings of criteria 4 (reliable data) and 5 (sound analysis). It is not always easy to rate criteria separately. For example, if the reliability of data is considered as poor (criterion 4), the assessor will hardly rate the analysis as sound (5) and the findings as robust (6).

Good or poor quality often depends on details, such as the use of a statistical source that suffers from a hidden bias, the use of an evaluation technique that is not totally appropriate or the implicit adoption of a causal assumption that is disputable. It calls for an experienced professional to detect these killer details. This is why several assessors are needed, including at least one expert of the policy domain under evaluation.

Unless criteria are rated as "excellent" or “unacceptable”, another difficulty arises from the need to sensibly take contextual constraints into account. Most often, the external evaluator is not in full control of the quality of the work. Other factors such as the resources allocated to the evaluation, deadlines, the quality of statistical data, the quality of monitoring information available, consensus on explanatory models, the openness of operational managers in providing information and conflicts of values are also influential. Therefore, assessors have to rate quality in relative terms rather than in absolute ones. A good report is what a good evaluation professional would have produced within the same set of constraints, and this cannot be assessed by someone other than a good evaluation professional. This is one of the reasons for which DG Agriculture hired an external expert to assist in quality assessment.
until the members of the evaluation unit built up their own capacity over a period of two years and a dozen assessments.

Who are good evaluation professionals and what do they produce within an average set of constraints? The least one can say is that international literature does not provide ready-to-use benchmarks of good quality. This is a reason why it is worth trying to draw some benchmarks from the practice of quality assessment in the European Commission.

**Indications about the current level of quality of European evaluators**

Considering the reports assessed to date, the evaluators have been generally good at answering the questions included in the terms of reference (criterion 1), at designing appropriate evaluation methods (criterion 3) and at gathering reliable data (criterion 4).

However these positive points must be seen in the light of a series of factors that are specific to the domain of the Common Agricultural Policy. First, these high ratings are linked to the fact that the evaluation unit invests much time in drafting precise evaluative questions and constantly insists on the need to answer these. Second, the quality of data collected has much to do with the comprehensive statistical databases and large number of research reports that are available in the domain of agriculture.

The reports that have been assessed are also satisfactory in the sense that they focus on results, outcomes and impacts, and go beyond management and implementation. However, this strength has to be nuanced because often the conclusions on impacts are purely
descriptive and fail to provide a cause-and-effect analysis. This is why criterion 5 (sound analysis) is one of the most poorly rated.

Whilst there have been some excellent reports, a series of weaknesses have also been repeatedly identified:

- judgement criteria and benchmarks are not made explicit enough;
- evaluators tend to accumulate an excessive amount of data, of which only a minor part is actually used to answer the evaluative questions;
- surprisingly, most conclusions are supported by one single source of information, although a more systematic cross-checking of different sources would often be possible and considerably enhance validity,
- conclusions are often not sufficiently grounded on evidence and recommendations are not derived from conclusions.

This picture draws on a limited number of twelve assessments. Does it indicate that the current state of quality of EU evaluations is weak? The evidence is not conclusive but several reasons tend to support such a generalization. First, the twelve evaluators originated from many different countries and backgrounds, which means the picture is not limited to a specific type of evaluator. Second, all evaluators had been selected through open calls for tenders. As selection is based, inter alia, upon quality criteria, it is probably fair to conclude that their work represents the best level of quality available on the market. Third, the picture is confirmed by other independent sources (Uusikyla & Virtanen, 2000; Toulemonde, 2001)
Through its repeated quality assessments, the DG Agriculture evaluation unit has progressively acquired some knowledge of the current state of the art. It cannot be excluded that this knowledge influences the scoring process itself. In fact, the intermediary scores ("good", "sufficient" or "poor") of the MEANS grid do not have explicit definitions, which means that assessors inevitably decide on these scores on the basis of their view of the state of the art, at least implicitly. For instance, a report which is scored as "sufficient" today will probably be judged more severely when the state of the art has improved.

**Credibility of quality assessment**

The quality assurance system puts a heavy pressure on external evaluators, something which has been surprisingly well accepted. Independently interviewed, some of these evaluators stated that they were satisfied with the clearer rules of the game. Quality criteria were generally accepted as legitimate.

Quality assessment of an evaluation report is not a purely technical operation, and it may in some cases become a sensitive issue. Since the Common Agricultural Policy involves large expenditure and high political stakes, evaluation reports usually attract the attention of various interest groups, lobbyists and the media. Conclusions of evaluation reports can always be used selectively to support the interests of particular groups, or contested where the evaluation does not conclude in their favor. In such debates also the quality assessment may become an object of attention. Depending on (1) whether the report concluded positively or negatively about the policy under evaluation and (2) whether the quality of the report was
assessed as good or poor, different arguments as to the credibility of the assessment may arise.

There have been cases where the quality of a report has been assessed “poor” while the report concluded negatively on the policy under evaluation. In such cases, DG Agriculture has been exposed to criticism like: “they rated the quality of the report poorly because they did not like the conclusions”.

In practice, DG Agriculture has repeatedly made positive quality assessments on reports that presented critical findings on the measures it operates. Some of these evaluation reports have led to significant policy reform proposals. Nevertheless the simple fact that such criticism arise casts doubts on the credibility of the quality assessment process. Its objectivity may be put in question because it is led by DG Agriculture itself.

This weakness may not be a killing one. First, because there are some checks and balances within the assessment process. In fact, the quality statement is discussed and validated within an evaluation steering group which typically comprises several members of other DGs that have vested interests in policy reforms. Second, this practice can be defended by the assumption that public managers are more willing to use evaluation findings when been personally involved in validating their quality.

Finally, the credibility of the first layer of quality assurance might be strengthened by a second layer that the European Commission has tried to develop and which will be described in the next section.
Assuring the quality of evaluation management

Setting up criteria: the Good Practice Guidelines

While the first layer of assurance is concerned with the quality of individual evaluation reports, the second layer focuses on the management systems that operate the evaluation process. The development of principles guiding the organization, programming, steering and reporting on evaluation activities has been part of the Commission’s efforts to establish an evaluation culture in the context of the administrative reforms. The idea of a “system level” control of evaluation quality dates back to the 1996 reform program (European Commission, 1996) which, among other things, established operating bodies and responsibilities for putting into practice the new culture of regular evaluation. These include:

- Setting up a network of officials responsible for evaluation in the different Commission departments.
- Defining explicit rules for the regularity of evaluations.
- Setting up coordinating bodies in the Directorate General for Budgets (with responsibility for promoting good practice, methodological advice and training) and in the DG for Audit (responsible for reviewing the evaluation systems).
- Establishing common instruments such as the Commission’s evaluation program and an Annual Evaluation Review.

The development of system level instruments for quality assurance has been initiated and managed centrally by the horizontal departments, unlike the quality assurance of evaluation
reports which is managed by the operational departments according to their own needs.

From standards to principles and guidelines

The Network of Evaluators (European Commission 2000c) was set up as one of the efforts to promote good evaluation practice in the Commission. It gathers evaluation managers and members of evaluation units of all departments. The Network provides an opportunity:

- to discuss quality standards and questions of evaluation design and methodology;
- to discuss joint sectoral/thematic evaluations involving more than one service;
- to reflect on the Commission’s Annual Evaluation Programme;
- to contribute the Annual Evaluation Review;
- to discuss training and instructional needs in the services, and to comment on the quality and relevance of evaluation training programs, drawing on the knowledge of those services with particular experience in this area; and
- to build on the experience of its members through constructive peer review of evaluation activities undertaken in the Commission.

The idea of developing formalised standards for the quality of evaluation was raised early on in the discussions of the Evaluation Network. In 1998 a working group was set up for this purpose. The DG for Audit was particularly proactive because it had been given the task of regular review of the organisations and systems for evaluation (European Commission 1996).

In the course of working group discussions the idea of obligatory standards was turned into
good practice guidelines, which would be a “softer” instrument for assessing progress in evaluation. The guidelines should reflect the state of the art in the Commission, and be updated as practices evolved. This approach was preferred to obligatory standards, which would have put unrealistic demands on those DGs that had not yet got very far in setting up evaluation systems.

The working group first took stock of regulatory frameworks and principles that had previously been established by the Commission. Interesting and innovative practices from some DGs were integrated. The work also drew from internationally available evaluation standards, in particular those of the OECD (OECD 1998) and the Australasian Evaluation Society (Australasian,1997) and the American Evaluation Association (Programme Evaluation Standards 1994). Consistency was also sought with evaluation guides issued by those DGs that already had established evaluation practices, such as the guide "Evaluating EU Expenditure programmes by DG Budget (European Commission 1997), the MEANS guides on evaluating socio-economic programmes by DG Regional Policy (European Commission, 1999a) and the guide on "Project Cycle Management" by DG Development (European Commission 1993).

The results of this work were formalised as the “Good Practice Guidelines for the Management of the Commission’s Evaluation Function”, the first version of which was formally adopted by the Commission in February 1999. As originally planned, they were subsequently updated and the version currently in force was adopted in January 2000 (European Commission 2000c).
The Guidelines are structured in a small set of general principles, and for each of them some concrete guidelines on how the principles should be put into practice. The principles cover the following four areas

- integration of the evaluation function in the overall management of the DG
- resources to be allocated to evaluation
- quality of evaluation
- external communication of evaluation results.

The focus of the guidelines is on the organization and management of evaluation within a DG. They are clearly geared to enhancing a managerial use of evaluation within the Commission for the purpose of policy design and implementation. The principles and guidelines for each of the four areas are summarized below.

**Principle 1: Integration of evaluation in the overall management**

The general principles in this area require that the management of the DG openly supports the evaluation function, and that the evaluation function is managed so that it reflects and supports the priorities of the DG and its operational management. The guidelines for ensuring this mutually supporting existence of evaluation and management include, for example, the following:

- The role of the evaluation function should be visible and the organisation chart should indicate clearly the official, sector or unit responsible for evaluation matters (the evaluation function).
• Due to its role in policy formulation and planning, the evaluation function should be clearly distinguished from the functions of control, audit and inspection and should have close links with unit(s) in charge of policy design.

• The Director-general should give a clear mandate to the evaluation function. This should include, inter alia, right of initiative as well as responsibility to design, monitor, validate, and disseminate evaluation results.

• An annual evaluation plan should anticipate future policy priorities and operational needs and should be approved by the Director-general.

• There should be feedback mechanisms appropriate for communicating effectively to management and relevant stakeholders all types of evaluation results. They should include a means whereby management verifies whether evaluation results are taken into account in programme management and/or policy proposals.

**Principle 2: Adequate resources**

The general principle put forward in this section states that the evaluation function should have sufficient and appropriate resources in terms of funds, staff and skills to discharge its responsibilities in a cost-effective manner. This clearly reflects the fact that the guidelines were prepared by a group of officials in charge of evaluation functions, also serving the purpose of defending their case in the internal competition for scarce resources. The guidelines for materialising the principle of adequate resourcing include:

• The DG should provide its evaluation function with appropriately skilled and experienced staff, and training.
Financial resources available for commissioning external evaluations should be clearly identified and planned.

The evaluation function should be given financial resources for conducting work on methodological issues.

Principle 3: Quality of evaluation

The principle covering this area stipulates that “Evaluation processes and reports should be governed by standards designed to ensure their quality, notably as regards relevance, rigour and timeliness”. This can be interpreted as a systemic approach to quality assurance. The systemic approach is further emphasised in the first concrete guideline to be followed: “Each evaluation function should establish, or refer to known quality standards. Standards may address both the process of evaluation and the evaluation results.”

However, the quality principle also include substantive quality requirements as can be seen from the following list:

- Evaluations should address questions and make recommendations which are realistic and relevant to the intended audience (substantive).
- The evaluation function should facilitate the evaluator’s access to the data sources (systemic).
- Stakeholders in the programme/action to be evaluated should be identified and their involvement in the evaluation process clearly stated (systemic).
- The final report should present findings, which are based on carefully described
assumptions, data analysis and rationale (substantive).

- The final report should present the findings determined by the (external) evaluator, and these should not be amended by without the evaluator’s consent (systemic).

- The evaluation function should communicate the quality requirements to the evaluators and the participants/stakeholders in the evaluation, support the application of these and develop some form of quality assessment (systemic).

The above practical advices are presented in the guideline as examples of standards that the DGs could include in their own evaluation quality assurance process. The MEANS-grid described above is the most well known of the standards referred to in the guidelines, and it is explicitly mentioned in the reference list of the document. However, some DGs questioned the applicability of this tool to their policy area. As a result the guidelines were formulated in a neutral way, while inciting DGs to develop their own quality assurance instruments, suitable for the particular features of their evaluation processes. At the time of the writing, DG Agriculture and DG Research were among the most advanced initiatives. As described in the previous section, the first one built upon the MEANS grid, while the second developed its own standards.

**Principle 4: Communication of evaluation findings**

The principle governing this area is that “as a general rule, evaluation results should be made publicly available”. Wide access to evaluation information is strongly supported by the guidelines. Also a focus on the utilization of evaluation is clearly one of the guiding principles of the practical advice put forward in this section of the guidelines:
• Evaluation results should be communicated in an appropriate form to meet the needs of decision-makers and European citizens as well as to fulfil reporting obligations included in the legal bases.

• Summarised information on evaluation results and their planned use is compiled annually into the Commission’s Annual Evaluation Review which is made available to the public through the Europa website. DGs should develop specific evaluation sites on which individual evaluation reports are available.

• It should be made clear whether a published report represents the views of the Commission or solely those of an external consultant.

• Commission services should ensure appropriate resources for planning and managing communication and include the communication of evaluation results within the DG’s overall communication plan.

• Information disseminated on the basis of evaluation reports should protect confidentiality as appropriate.

The Good Practice Guidelines are still valid and referred to in various Commission documents. Also the Court of Auditors has referred to this document in developing its own approach to assuring the quality of information, as will be described below.

In the context of its current administrative reform the Commission issued a new policy document on evaluation (European Commission 2000a), where it instructs the DGs to follow the general principles expressed in the existing good practice guidelines while also acknowledging that the guidelines need to be further developed. Need for further precision is signalled as regards the division of tasks between evaluation and the newly reinforced audit
functions of the Commission. The same policy document also states that in addition to the guidelines on evaluation management there should be unified standards for the quality of evaluation reports. At the time of writing, the work for further development of the guidelines and the establishment of unified quality standards was still in progress.

**Auditing evaluation systems**

**Experiment with central systems audit: 1997-99**

The idea of a systems-based approach to the control of evaluation quality was present although not yet fully developed in the 1996 policy document (European Commission, 1996) defining the steps to be taken to establish an “evaluation culture” to the Commission. Formally the task entrusted to the DG Audit was described as “a regular review of the organization and systems DGs have put in place to achieve proper evaluation and feedback”. In practice this took the form of an annual report to the Commission, in which the quality of each DGs evaluation systems were assessed against the Good Practice Guidelines.

In these annual reports DG Audit classified all DGs’ evaluation systems as “mature”, “implemented” or “emerging” depending on how well their organisation and management of evaluations corresponded to the Guidelines. The reports also presented an overall assessment of progress made in developing evaluation systems. The reports themselves are internal documents, but summaries of the assessments have been included in communications that are publicly available (European Commission 1999 and 2000b). The 1998 report (EC 1999) concludes that “the organisation and systems for evaluation are well developed in three policy areas (Structural funds, research and technology policy and external relations), that good
progress has been made in the common agricultural policy and that also in the remaining areas some progress can be noted”. Need for improvement is reported in the use and quality of evaluation, and a particular critical observation concerns the fact that most departments have not defined any standards for their evaluation process or for the quality of evaluation reports.

The 1999 Report on the Commission’s Organisation and Systems for Evaluation summarises the situation in the following way (European Commission, 1999): “despite progress made in many departments, serious inadequacies remain to be remedied if evaluation is to measure up to its expected key role …. In particular:

- ex-ante evaluations for new programmes or actions are not carried out regularly
- evaluation findings and conclusions are not systematically applied in decision making
- quality control of monitoring and evaluation processes is poor.

Unless these weaknesses are addressed, strategic planning and operational programming will be made on inadequate informational basis. To ensure quality control at all stages of the evaluation process, the question of resources and skills should be addressed as well.”

This critical position reflects no doubt the general fierce criticism that was addressed at the Commission and its management systems after the dismissal of the Santer Commission, and the high expectations for the administrative reform for which the first blueprints existed at the time when this report was presented.
Including evaluation in the management audits

The 1999 report on organisation and systems for evaluation is the last one of these reviews. The DG Audit disappeared in the course of a Commission’s reorganisation and the embryonic systems-based audit remained as a short-lived experiment (1997 – 1999). In the future, the practice could be revived by the newly established Internal Audit Service (IAS), which is an independent internal body with a wide mandate to audit the management in all Commission departments. However, evaluation is no longer mentioned as a particular topic to be controlled. The IAS bases its work programme on risk assessments, taking into account also possible requests from the senior management of the Commission. So far evaluation has not been perceived as an area with particular risks and is not mentioned in its work programme for thematic audits, nor is it targeted in the IAS’s reform progress audits (European Commission 2001b and 2001c).

In addition to the central IAS, a new “internal audit capability” is now attached to each Director General with an aim to assure the quality of management, including evaluation. Both the DG level audit capabilities and the IAS base their work on the Internal Control Standards which the Commission adopted in 2001 (European Commission 2001a). These control standards cover the whole area of management from setting objectives and performance measurement to financial procedures and evaluation.

Evaluation is covered by one of the 23 standards, which is formulated as follows: “Each DG shall establish or have access to a properly staffed evaluation function responsible for carrying out or commissioning ex ante and ex post evaluation of all its activities. It shall prepare an evaluation plan which sets out the timing of the planned evaluations and against
which progress is regularly reviewed. It shall ensure the systematic follow up of the conclusions of evaluation reports.”

This, together with the 2000 policy document on evaluation (EC 2000a) shows that as the results of the administrative reform process unfold, the commitment to regular evaluation by the European Commission has been consolidated. However, it may seem that the instruments for quality assurance have not been reinforced. Overall management control systems have been modernised and become more systematic, but evaluation management is no longer a subject of a specific quality assurance scheme. The primary responsibility for audit of evaluation systems is in each DG’s internal audit capability, and in the second instance the centralised IAS may focus its audit activities on evaluation management. As the main risks in the Commission’s management have traditionally been perceived to be in other areas, for example in procurement procedures and financial operations, it is not very likely that evaluation would become a specific target for the internal control process.

Even if it seems that the formal quality assurance systems for evaluation have remained thinner than originally planned in the 1996 plans (European Commission, 1996) for establishing an evaluation culture, there are, nevertheless, practices that may at least indirectly serve as effective instruments for quality assurance of evaluation:

- The Evaluation Network provides a forum for exchange of experience and may constitute an implicit peer review.
- The principle of publishing evaluation reports will be progressively enforced, which will create an incentive for assuring quality.
• The Annual Evaluation Review prepared by DG Budget may also put pressure on the DGs to assure the quality of evaluation information although it is not meant to be a quality control instrument. Its main aim is to summarise and communicate evaluation information in a digested and accessible form to the decision makers. Comments on quality are included in this document to qualify the information presented. As the main audiences of this report, the Budgetary Control Committee of the Parliament and the Budget Committee of the Council are keen consumers of evaluation information, they may and do put pressure on the Commission to react to any quality deficiencies noted in a public document.

So far the trend seems to have been from formal and regular systems audit focused on evaluation towards a less formal approach combining a voluntary commitment to quality at DG level and systems audits addressing evaluation as just one among other management practices. Peer group control in a professional community might be an effective way of inciting evaluation managers to keep committed with relevance and reliability, especially if they risk public criticism in case of poor quality.

Another significant shift might affect the demand for evaluative information. For years, the Council and the Parliament have pressed the Commission to evaluate. Since this demand is increasingly satisfied, they might soon turn to requiring high quality evaluations. This shift is currently anticipated by the European Court of Auditors (see next section).
Assessing the quality of information provided to policy makers

Evaluation and the European Court of Auditors

Growing interest in evaluation

The European Court of Auditors (hereafter the Court) is the institution charged with the external audit of the EU budget. It undertakes traditional financial audits and “sound financial management” audits, which may be described elsewhere as “performance” or “value for money” audits. It reports on these to the European Parliament and Council of Ministers and thus assists them in holding the Commission to account for its management of the budget.

As well as fulfilling its role as defined in the Treaty, the Court aims to “assist and, where appropriate, advise those responsible for deciding upon and managing Community programmes so as to improve financial management” (European Court of Auditors, 2001a).

Over recent years, the European Court of Auditors has taken an increasing interest in the quality of the evaluations carried out by the Commission’s services. In doing so, it has several reasons:

• sound evaluation of outcomes is part of the overall management cycle and thus falls under the responsibility of the Commission;

• the Court does not itself have the resources to carry out comprehensive evaluations or to commission and manage these directly, so it increasingly seeks to rely upon those carried out by or for the Commission to enrich its sound financial audit reports;
• as increasing numbers of evaluation reports are sent by the Commission to the European Parliament and Council as part of the accountability relationship, it is increasingly important that the Court is able to assure the political authorities of their relevance and reliability.

This increased interest in evaluation is connected to the development of accountability relationships and to the recent thinking about governance (IFAC, 2001). The Court is aware of the evolving role of Supreme Audit Institutions which has already occurred in some countries and is beginning to gather pace elsewhere away from an essentially fault-finding, critical approach to an assurance-based value-added emphasis. In this respect, the Court is re-assessing its fundamental relationship with its auditees. A central part of this is the ongoing process of developing a model of information management in the European Union, with the aim of identifying explicitly the activities for which the Commission could reasonably be held accountable.

Integrating evaluation into a model of management information

The making of European policies is a complex process, intrinsically shared between the three “Institutions” (Council, Parliament and Commission). As a consequence, public administrators cannot be taken as responsible for the efficiency or inefficiency of policies which result from long negotiations between the three Institutions.

The model of information management enlarges the boundaries of their responsibility in terms of outcomes. In the traditional model, public managers are fully responsible for their processing of inputs (public expenditures and staff) into outputs (services and facilities supplied). On the other hand, they do not control and they cannot be held responsible for
outcomes like the resolution of environmental problems, the satisfaction of social needs or the proper handling of economic stakes.

To a certain extent, the model of information management extends the responsibility of public managers in the direction of outcomes. In this model, public administrators are responsible for gathering information about inputs, outputs and outcomes and need to take stock of that information relevant to the tasks that are under their responsibility, i.e. operational decisions and proposals for policy decisions.

This model introduces new concepts of inputs and outputs and makes public administrators responsible for them. Input is defined as the gathering of monitoring and evaluative information. Outputs are defined as operational decisions and policy proposals that should be supported by monitoring and evaluative information.

In fact, the new concepts include evaluative information, which in turn includes the traditional concept of outcomes. This means that the new model makes public administrators responsible for information about outcomes, if not for the outcomes themselves. The model helps to define more precisely the auditors’ activities. It clearly allows them and, increasingly, obliges them to assess information systems. The Court's approach in this area is an emerging one, based on the experience gained in a number of audits and, for the moment, only partially reflected in the manual that guides the auditors. One can, however, identify three categories of criteria that have been applied. The three categories are:
Category 1: availability of sound information

- Did public managers produce accurate, reliable and timely information about implementation, achievement of objective, actual impacts, expected impacts, economic and social needs, critical changes in the context?

- Did that information reflect the concerns of various stakeholders?

Answering these questions requires that the auditor goes beyond an examination of the direct evaluation process and also investigates the way in which that process was managed. As has been seen above, important advances have been made within the Commission (both through the use of the MEANS grid and as a result of the Good Practice Guidelines) to assure the quality of the evaluation process.

A next step for the Court will be to develop a systems-based methodology which it can use to assess firstly whether this quality assurance system is, in principle, effective in producing high-quality evaluations and whether the system has been properly applied in the case of the specific evaluation that is the subject of the audit.

However, it is rather more difficult to establish objective criteria to assess other aspects of the management of the evaluation process. For example, the auditor needs to arrive at a conclusion about the quality of the questions that the evaluators have been asked to answer. Are they the most relevant questions? Are their scope and range appropriate?

The difficulty of drawing up and applying objective and meaningful audit criteria in this key area may explain why the Court does not yet regularly carry out audits of evaluations and the evaluation process. Audit criteria are usually based upon legislative requirements,
professional standards, best practice or some other form of generally-accepted norm against which the activity being audited can be compared and objectively assessed.

In the area of setting the questions that an evaluation is to answer, it is possible to imagine norms that might eventually find their way into professional standards for evaluators or into the Commission’s Good Practice Guidelines. For example, a recommendation that stakeholders should be involved in the evaluation process already provides some additional assurance as to the quality of the questions, and thus would provide a starting point from which the Court might develop the necessary audit criteria.

**Category 2: use made of the information**

- Was that information presented to decision makers in a way which suited their needs and absorption capacity?
- Did public managers consider and discuss reforms, alternatives or policy options that derive logically from the available information?

The item above is perhaps common for all policy-related information presented to decision-makers. There is always a need for synthesis, but this must not result in biased or inaccurate communication. This issue is already dealt with by the ninth criterion of the MEANS grid, discussed above. However, in the EU context, there is also a linguistic dimension: whilst the Commission typically works in English or French, the European Parliament and Council of Ministers operate to a much greater extent in more than ten official languages. There is thus an additional challenge to assure the quality of the translations of these synthesis reports.
Criterion 3: decisions and actions taken

- Did operational and policy decisions refer to the available information and explain how it has been used?

**Box 1: Auditing information systems: example of the EU Sugar Policy**

... The Commission's management of the [Sugar policy] is assessed in terms of inputs (information) and outputs (regulations, decisions and proposals). The main Commission activities are covered, i.e. the proposals for production quotas, the annual proposals for institutional prices, the decisions on the export programme and the financing system. The extent to which the Commission has assurance that the [Sugar policy] is functioning well is also examined, along with how the Commission responds to events affecting the [Sugar policy] and to the different stakeholders

... The quota arrangements, a central element of the [Sugar policy], were introduced in 1968 for a transitional period and subsequently renewed on several occasions and most recently in 1994 for the period 1995-2001

It is clear from the above that the Commission did not present all the appropriate information to justify the continuation of the quota arrangements for 1995 to 2001, nor did it give a comprehensive analysis of the state of achievement of the [policy] objectives. At the same time it was noted that extensive, although not complete, information was available to the Commission services. Furthermore, the Commission did not present the likely impact of their proposals compared with other options and there was no assessment of the consequences of the regime for the consumer, nor of the impact on the environment

*Extracts from the Court’s report available on the Europa website*

The example in box 3 is one of the early attempts within the Court of Auditors to apply the new information management model. It deals with policy decisions taken in 1994, i.e. largely before evaluation was introduced in the DG Agriculture, although the Court's report was published in February 2001, at about the same time as the evaluation quoted in box 1.

To briefly cite another example, in its report on the management by the Commission of the program of assistance to the Palestinian society (European Court of Auditors, 2001b) the
Court sought to define those areas over which the Commission should have been able, by obtaining and analyzing reliable and timely information, to exert management control and influence over outcomes in a highly volatile situation, and to assess the reactivity of the Commission’s management to that volatility. This report focuses strongly upon the management information model, described above, that is evolving in the Court. Incidentally, the report also refers to an independent evaluation of the main projects within this program but, as is understandable given the “state of the art” at the time the audit was carried out, did not seek to assess in the systematic way described above the quality of that evaluation.

The above illustrates that there have been developments within the Court and it can be noted that there is a significant degree of interest amongst its auditors in the evaluation process and their role in it. The quality and use of information by managers and policy-makers is an issue that is increasingly addressed by the Court's audits. In the guidance issued to auditors, this is a matter that figures strongly in the performance audit planning process. However, it cannot yet be said that the Court is playing a systematic role in assessing the quality of evaluative information. Furthermore, the new accountability relationships described above are only beginning to emerge now in the European institutions.

Towards an integrated system of quality assurance

It is almost by random that it has been possible to illustrate two parts of this chapter with examples relating to the same ‘Sugar Policy’ (see boxes 1 and 2). In fact both exercises were independent and the evaluation itself cannot be seen as a response to the Court’s remarks. If conclusions were to be drawn from such co-incidence, then the same policy should be subject
to:

- a DG assessing the quality of the external evaluation of a given policy or program,
- some internal audit verifying that this DG manages its evaluations properly,
- an external audit verifying that decision-makers in this policy area have actually received and used proper evaluative information.

When considering the current development of evaluation and audit practices in the EU Institutions, it can be expected that such three-layered examples will be available within the next few years. So as to summarize this chapter, let us come back to the three layers as follows:

- The first layer of quality assessment applies to external evaluators’ reports. It is handled by the evaluation managers (or evaluation function) in line departments (DGs). This assessment relies upon an quality assurance system which is internal to the public organization. The system is both compatible with internationally recognised standards and finely tuned to the specific needs of the policy area. At this level, evaluative information is assessed in both a formative and summative way, each evaluation report being assessed mainly in terms of substance.

- The second layer of quality assessment applies to the evaluation system which is run in the first layer. Therefore it looks like systems audit but it may range from a formal audit to an informal peer review. It may be operated in a centralized or decentralized way or both. Some internationally agreed standard also apply at this level, as can be shown by the fact that the EU Good Practice Guidelines is well accepted. Quality assessment at this level
may be either formative or summative. In the later case, it is run as a formal audit and used to strengthen the credibility of the first level of quality assessment, something which will be increasingly needed as long as evaluations become subject to public debates.

- To a large extent the two first layers consider quality with the standpoint of the producers of evaluative information, since they apply standards resulting from a consensus among experts. On the contrary, the third layer applies to the whole set of evaluative information received by decision-makers, with their standpoint, that is to say with the consumer’s standpoint. Quality standards at this level should mainly derive from users’ needs and not from the evaluation profession. At this third level, external audits have to question the value added of evaluation in order to ensure that the two previous layers of quality assurance remain user-oriented (Pollitt & Bouckaert, 1995).

The reader might consider that this three layered quality assurance system would be a decisive step towards an “evaluation explosion”, echoing the audit explosion that Porter () has described, with similar undesired consequences. This would be a good subject for a new chapter, but the practice of evaluation quality assurance has to develop first before its unintended consequences can be judged positively or negatively.

In the short run, this chapter has raised practical arguments in favour of a multi-layered approach:

- The quality of an individual evaluation report cannot be properly evaluated without the help of programme managers because only them can discover biases and errors in highly technical areas. However, their involvement will inevitably cast doubts on the impartiality
of the assessment. Adding another layer of quality assessment is a reasonable means for legitimizing the first one.

- The first layer(s) of quality assessment inevitably rest in the hands of evaluation professionals. This calls for another layer more clearly geared at assuring quality with the standpoint of the users.

- Not all layers need to be structured in a rigid way, nor all assessments at a given level. Formative assessments and self assessment under the eye of peers may be sufficient to ensure professional competence and ethical professional conduct.

References


European Commission. 1996. SEC 96/659 final - *Communication to the Commission by Mr Liikanen and Mrs Gradin, in agreement with the President. Evaluation : Concrete steps towards best practice across the Commission*. Brussels: EC


European Commission. 1999b. SEC(1999)69/4 - _Communication from Mrs Gradin and Mr Liikanen in agreement with the President. Spending more wisely: Implementation of the Commission’s evaluation policy_.

European Commission. 2000a. SEC (2000)1051 - _Communication to the Commission from Mrs Schreyer in agreement with Mr Kinnock and the President. Focus on Results: Strengthening evaluation of Commission activities_. Brussels: EC.


European Commission. 2001a. _Internal Control Standards_. Brussels: EC.

European Commission. 2001b. _Communication to the Commission by Vice-President Kinnock in agreement with Commissioner Schreyer. Internal Audit Report for 2000_. Brussels: EC.

European Court of Auditors. 2001a. *Court Audit Policies and Standards*. Available at www.eca.eu.int


---

**Notes**

i In a subsequent reform this DG Audit control function has been suppressed, and an additional coordination function has been set up in the Secretariat General of the Commission. This new central coordination body is responsible for policy evaluation in the context of the Commission’s strategic planning process.

ii “Evaluation function” refers to the organizational entity that has particular responsibilities for carrying out or coordinating evaluation activities in a DG. The Commission documents discussing evaluation use this term because the type, level and size of these entities vary...
across DGs. Since DGs are left to decide whether they designate a unit, a sector or only an official for this purpose, a neutral term has been sought to refer to any of these.