

Jean-Claude Barbier **Bernard Perret**
Directeur de recherche (CNRS) Chargé de mission au Conseil Général des
Ponts et chaussées
Centre d'études de l'emploi/Université Ministère de l'équipement, des transports
et du logement
Paris VII – Denis Diderot Membre du Conseil national de l'évaluation
Secrétaire de la Société Tour Pascal B
Française de l'Évaluation 92055 La défense cédex
Centre d'études de l'emploi France
29, Promenade Michel Tel 33 (0) 1 40 81 60 31
Simon Fax 33 (0) 1 40 81 23 24
93166 Noisy le Grand Cedex Bernard.perret@equipement.gouv.fr
France
tel: 33 (0)1 45 92 68 54
fax: 33 (0)1 49 31 02 44
Jean-Claude.Barbier@cee.enpc.fr

Ethical Guidelines, Process and Product Quality Standards, What For?

An SFE (French Evaluation Society) Perspective

Paper presented at the European Evaluation Society Conference
Lausanne,
October, 12th -14th 2000

Introduction

Since its rather recent foundation in Marseilles (June 1999) the French Evaluation Society (SFE) has harboured a special working group on “Standards and Ethics”¹. To the group’s members, it seemed logical to start their investigations by turning to practice elsewhere. Other Evaluation Societies (ES) in the world provided us with important food for thought and material². The societies’ diverse origins, history and societal contexts seem to account for the considerable variety of their approaches to the question. And this is true despite a clear and overwhelming influence exerted by the Joint Committee on Standards for Educational Evaluation (JCSEE) reference (1994 version)³.

This paper will first present a brief analysis of the main purposes and contents of ES guidelines and standards⁴, as we see them from a French perspective. We then turn to explaining why the group chose to propose SFE members with a progressive and still tentative path to establishing references for a wide variety of French evaluation circles⁵. In the paper we propose to deal separately with four types of norms, i.e. norms addressing quality (either of products or of processes), norms concerning ethical conduct and, finally, norms concerning the social value of evaluation. We especially endeavour to differentiate between these norms, whether standards or guidelines, on the basis of their specific potential use.

¹ Its French name is « *Standards et déontologie* ». Ethics is used here as an equivalent of *déontologie*, despite differences between both notions. *Déontologie* and *Propriety* are also close. Standards featured as a plenary session theme at the second SFE conference in Rennes in June 2000. The views presented here have been discussed at the Rennes meeting; while they broadly mirror the working group’s orientations, they are the authors’ and they have not been yet formally endorsed as SFE policy.

² J.C. Barbier wishes to particularly thank Chris Milne of the Australasian Evaluation Society and Thomas Widmer of the Swiss SEVAL, for their help.

³ See next sections.

⁴ ES societies considered here are the following: the Australasian Evaluation Society, the American Evaluation Association, the Canadian Evaluation Society (Société canadienne d’évaluation), Associazione Italiana di Valutazione, SEVAL (Suisse), DEGEVAL (Germany), Société wallonne d’évaluation et de prospective, and the UK Evaluation Society.

⁵ Contrary to some other countries it would be an exaggeration to describe France as having an “evaluation community” as such.

I – Standards versus Ethics?

Ethics and Quality

Most ES surveyed distinguish principles of *personal ethics* (personal standards, guidelines for an ethical conduct), and *quality standards*. However this distinction is not made along homogeneous lines - we will return to that in the next sections. We thus have to forsake the quest for a clear-cut separation between *quality* that would apply to evaluation processes and products/outputs as well as methods on the one hand, and *ethics* that would only be valid in the realm of personal responsibilities on the other.

C. H. Weiss in her classic book (1998, p. 109-112), is certainly among those who insist upon links between evaluation and politics. She focuses on the evaluator's behaviour: "*the evaluator needs to keep standards of ethical behavior in mind*" (1998, p. 109)⁶. She deals with ethics in various chapters: concerning evaluation planning, she for instance states that ethical issues "*deserve high priority*" and recommends five "*principles*", i.e. "*honesty, informed consent, confidentiality and anonymity, high competence and reciprocity*". All five are seen here as ethical imperatives and she relates them explicitly to some of the American Evaluation Association (AEA) Guidelines (p. 92-95). She also applies ethical conduct to collecting data (p. 175). But her main developments are under chapter 5 ("Roles for the evaluator", p. 109-112) where she states that evaluators should protect staff and client interests. She endorses the JCSEE rule for dealing with conflicts of interests and advocates "*openness of communication*" and "*candor*". At the end of her book, she writes that all her advice on matters of ethics "*can be condensed into two rules. Do not harm the people studied, and do not distort the data*" (p. 325), with the caveat that "*simple as these rules seem, they can collide with each other*". The "*highest quality*" of the evaluation study is thus seen as the main ethical imperative that condenses all the rules.

Patton (1997, p. 16) insists upon the fact that not only questions of technical quality and methodological rigour are at stake. He therefore supports the four JCSEE guidelines, i.e. "*Utility, Feasibility, Propriety and Accuracy*"⁷ (where ethical imperatives mainly appear under the "Propriety" item). When dealing with the political content of evaluation (in his Ch. 14, "*Power, Politics and Ethics*"), he also insists on the close relationship between politics and evaluation (p 341 ssq). For him, ethical aspects of "*utilization focused evaluation*" mainly concern two points, (i) limiting stakeholder involvement to primary intended users and (ii) working closely with the users. To him, this amounts to the main ethical question: "*Who does an evaluation – and an evaluator, serve?*". Should the evaluator mainly work for his clients and to what extent does he or she has to take into account "the rest of society" and go "*beyond the immediate welfare of the immediate client*"? These questions clearly remain controversial and certainly cannot exhaustively be dealt with via standards. However it remains a central question for which all of us evaluators have to choose their own way and

⁶ This applies in a context where the evaluator is not only accountable to the client but responsible to a much wider constituency: "although some uses of the study are immediate and obvious, much use will be conceptual and difficult to discern with the naked eye (...) the evaluator has to shape the evaluation to answer important questions, so that what the varied audiences learn will lead to policy and program improvements"(Weiss, 1998, p. 45).

⁷ The detailed content of each of these is discussed in part II.

clearly state it. One solution is to argue that because evaluation is always political in its content, the evaluator always is accountable to the general public (Barbier, 1999) or to society as a whole (Conseil Scientifique de l'Évaluation, 1996).

One substantial difficulty is how to conciliate the legitimate interests of all groups and persons affected by a particular evaluation and a more global responsibility towards general interest. Even taking for granted that evaluation and control (or audits) are clearly separated from each other, it is quite unavoidable that, when carried from the point of view of general (or public) interest, evaluations will lead to conclusions and recommendations that might harm some groups' interests or reputation. We have no definite answer to the question as to whether it is possible and desirable to fix limits *ex-ante* to individual consequences, or individual accountability resulting from a particular evaluation.

This dilemma clearly appears when confronting items 4 (respect for people) and 5 (responsibilities for general and public welfare) of the AEA Guiding principles:

AEA Guiding Principles for Evaluators (general definition)

- **Systematic Inquiry:** Evaluators conduct systematic, data-based inquiries about whatever is being evaluated
- **Competence:** Evaluators provide competent performance to stakeholders
- **Integrity/Honesty:** Evaluators ensure the honesty and integrity of the entire evaluation process
- **Respect for People:** Evaluators respect the security, dignity and self-worth of the respondents, program participants, clients, and other stakeholders with whom they interact
- **Responsibilities for General and Public Welfare:** Evaluators articulate and take into account the diversity of interests and values that may be related to the general and public welfare

Seen in the French context of evaluation, one might assume that the dilemma is related to a dialectical relationship between the basic conditions for the possibility of evaluation as a distinctive activity on the one hand, and the evaluators' accountability. Standards and guidelines, either from a quality or ethics point of view may render this relationship possible. On one hand, following Leca (1997, p. 11) we need an "*area of autonomy*" for evaluation, separate from the rest of the political and managerial systems. But to foster it, it is not possible to rely *only* upon evaluators' personal ethics; collective norms have to be agreed upon that support and preserve the evaluators' independence, especially against the risk of their cynical and strategic instrumental use by politicians or managers. On the other hand, it seems thoroughly illegitimate that evaluators claim their independence without at the same time actually abiding by collective and public criteria to assess the quality of the job they achieve. These criteria's substance should be publicly available to all possible stakeholders.

Evaluation Societies references: an overview

Three main conclusions stem out of our *partial* review of other ES references known to us. Certainly this survey could not claim any exhaustiveness. Moreover we think it is very important to stress that a crucial factor to the comparison of standards across the world is a rational assessment of how these are actually *implemented* and possibly linked to sanctions. We lack most of this knowledge, despite a few insights gained from exchanges with our correspondents.

The first conclusion is that two main types of norms exist, Standards and Guidelines. To our knowledge, only the Australasian society seems to work on a third type of normative

reference, a “Code of Ethics”⁸. But, as the table below shows, variety across societies is the rule.

Countries/Regions	Societies	Standards	Ethical Guidelines/ Guiding Principles
Switzerland	Société suisse d'évaluation (SEVAL)	JCSEE inspired Standards	None
Germany	Deutsche Gesellschaft für Evaluation (DEGEVAL)	Working group on JCSEE/Swiss standards	None
United States	American Evaluation Association (AEA)	JCSEE standards	Guiding principles for evaluators
Australia and New Zealand	Australasian Evaluation Society (AES)	JCSEE standards	- Guidelines on ethical conduct of evaluation - Working group on a Code of ethics
Italy	Associazione italiana di valutazione (AIV)	None	Linea guida per un codice deontologico del valutatore
Canada	Canadian Evaluation Society (CES-SCÉ)	JCSEE inspired Standards	Guidelines for Ethical Conduct
Wallonie	Société wallonne d'évaluation et de prospective	Working group	
United Kingdom	United Kingdom Evaluation Society (UKES)	Working group (?)	
France	Société française de l'évaluation (SFE)	Working group	

Different types of actors participating in evaluations may use standards or principles. However, only the Australasian Society clearly states that “guidelines are directed to people in Australia and New Zealand who commission, prepare, conduct and use evaluations, as well as those who research, teach and publish about evaluation”. All other references surveyed seemed to focus on evaluators as accountable in “the last resort”.

Our second conclusion is that, apart from the fact that guidelines are overwhelmingly devised for evaluators, there is no clear distinction between the substantive content involved in Standards on one hand and in Guidelines on the other. Testimony of the overlapping nature of both types of references may be illustrated by the comparison between the Guidelines surveyed and the JCSEE “Propriety” items. The Australasian Guidelines for Ethical Conduct, the Canadian Guidelines for Ethical Conduct and the American Guiding principles for evaluators, all seem to be centred on the individuals’ ethics but the “Propriety” standards also entail an ethical dimension:

⁸ This code would entail more institutionalization of the ethical guidelines and imply possible sanctions for the society’s members when they break the rules or harm the society’s reputation. For its part, the Canadian society is studying possibilities of engaging into certification activities.

JCSEE “Propriety Standards”

The propriety standards are intended to ensure that an evaluation will be conducted legally, ethically, and with due regard for the welfare of those involved in the evaluation, as well as those affected by its results;

- P1 Service Orientation -- Evaluations should be designed to assist organizations to address and effectively serve the needs of the full range of targeted participants.
- P2 Formal Agreements -- Obligations of the formal parties to an evaluation (what is to be done, how, by whom, when) should be agreed to in writing, so that these parties are obligated to adhere to all conditions of the agreement or formally to renegotiate it.
- P3 Rights of Human Subjects -- Evaluations should be designed and conducted to respect and protect the rights and welfare of human subjects.
- P4 Human Interactions -- Evaluators should respect human dignity and worth in their interactions with other persons associated with an evaluation, so that participants are not threatened or harmed.
- P5 Complete and Fair Assessment -- The evaluation should be complete and fair in its examination and recording of strengths and weaknesses of the program being evaluated, so that strengths can be built upon and problem areas addressed.
- P6 Disclosure of Findings -- The formal parties to an evaluation should ensure that the full set of evaluation findings along with pertinent limitations are made accessible to the persons affected by the evaluation, and any others with expressed legal rights to receive the results.
- P7 Conflict of Interest -- Conflict of interest should be dealt with openly and honestly, so that it does not compromise the evaluation processes and results.
- P8 Fiscal Responsibility -- The evaluator's allocation and expenditure of resources should reflect sound accountability procedures and otherwise be prudent and ethically responsible, so that expenditures are accounted for and appropriate.

Only some of these standards are clearly linked to personal ethics (“rights of human subjects”, “human interactions”, “fiscal responsibility”, “conflicts of interest”). Others rather refer to procedural requirements (“formal agreements”, “disclosure of findings”), quality of the product (“complete and fair assessment”) or even to the social value of evaluation (“service orientation”). On the other hand, “Propriety” standards do not address the issue of evaluator’s skills and competence, which is dealt with by the JCSEE via one the “Utility” items. (“U2 Evaluator Credibility -- The persons conducting the evaluation should be both trustworthy and competent to perform the evaluation, so that the evaluation findings achieve maximum credibility and acceptance”).

But the variety of guidelines also appears when examining the way they are structured and hierarchically organized, as is shown from a comparison between American and Canadian guidelines (following tables).

I - Common principles (competence and integrity)

Principles	CSE-SCÉ	AEA
	Guidelines for Ethical Conduct	Guiding principles for evaluators
Competence	Evaluators are to be competent in their provision of services	Evaluators provide competent performance to stakeholders
Integrity	Evaluators are to act with integrity in their relationships with all stakeholders.	Integrity/Honesty Evaluators ensure the honesty and integrity of the entire evaluation process

II –The third Canadian principle

Accountability	Evaluators are to be accountable for their performance and their product
-----------------------	--

III – The three other American principles

Systematic inquiry	Evaluators conduct systematic, data-based inquiries about whatever is being evaluated
Respect for people	Evaluators respect the security, dignity and self-worth of the respondents, program participants, clients, and other stakeholders with whom they interact
Responsibilities for General and Public Welfare	Evaluators articulate and take into account the diversity of interests and values that may be related to the general and public welfare

All in all, these comparisons lead to the following observations:

- (i) Guidelines are more in relationship to ethics and they lead to assessing personal conducts of people involved in evaluations; in some cases they might be used for sanctioning those who do not abide by them;
- (ii) An explicit judgement of conduct generally applies only to evaluators among evaluation actors;
- (iii) Issues of quality are closely intermingled with ethical issues;
- (iv) The “general/public” interest is not homogeneously addressed (see next paragraphs);
- (v) Sets of standards and guidelines leave all options open as to how they might be used (subscribing to, complying with, leading to certification, leading to professional sanctions, etc.)

Our third conclusion, which ought to be analysed much more in-depth, refers to the different countries' idiosyncrasies. Some of these are summarily presented in the following table:

Specific Items	Australasia	Switzerland	United States	Canada	Italy
Main significant specificity	Norms apply to all actors A possible Code of Ethics	Pragmatic approach (<i>Praxisnahe</i>) Implementation has just started	Where the standards originated Long experience	A certification hypothesis	An evaluator-centred approach (obligations/ <i>doveri</i>)
The general/public interest and protection of persons issues	Addressed in guidelines (risks to the clients, conflicts of interests, inequalities among stakeholders, informed consent, confidentiality, etc.)	Non explicitly addressed Propriety standards deal with: Conflicts of interests, mutual respect, human interactions, etc.	Guiding principles: respect for people and for public welfare Propriety standards	Non directly addressed in Guidelines Propriety standards	Non explicitly addressed in the obligations Most propriety standards are addressed
Other items of interest stressed	Honourable competition; Report significant problems; Report fully reflecting findings (possibly tailored to a given stakeholder group)	Timeliness, dissemination, cost effectiveness	Timeliness, dissemination, cost effectiveness Relationships with other professional standards	Timeliness, dissemination, cost effectiveness	Relationships with other professional standards

Of crucial importance to our subject is also the wording of guidelines and standards and the limited survey we have conducted by now show that many problems arise from translation. Obviously it is not desirable that each national or regional community of evaluators stick to their particular set of terms and claim that their wording are strictly dependent on the various national languages⁹; there are certainly many universal notions in evaluation across linguistic borders. Nevertheless, there is much more at stake to this issue than simply finding lexical equivalents.

⁹ Although minor, an interesting instance of this may be taken from the Swiss standards, which are translated both in German and French with sometimes slightly different and not exactly equivalent terms. In the "Feasibility" section, item two of the Swiss standards is equivalent to the JCSEE "F2 – Political viability" standard. It says in English: "evaluations should to be planned and conducted with anticipation of the different positions of various interest groups". The French wording is: "en tenant compte des positions différentes des groupes d'intérêt en cause" (in consideration with the different positions) and seems fairly equivalent to the English text. However the German version writes "unter Berücksichtigung der unterschiedlichen Positionen der verschiedenen Interessengruppen" where *Berücksichtigung* (consideration) bears multiple undertones as for what type of consideration it means.

That there is no exact French equivalent for “accountability” is well known and that our Canadian colleagues have translated it “*imputabilité*” in their guidelines does not solve the problem. Because the absence of a French equivalent is to be explained in “societal coherence” terms, i.e., taking into consideration the particular type of relationships existing between State and civil society in France. Another instance is “welfare” with its wide variety of meanings including happiness, well being, good fortune and health, but also prosperity, etc., not to mention the underlying reference to welfare states. There is a wide gap between the French “*intérêt général*” and “public welfare”, which certainly should not be underestimated when the debate about ethics is at stake. Other difficult candidates to translating are “honesty”, “propriety”, “integrity”, etc. These observations point to the very important caveat the AEA gives in its presentation of its principles, namely that “these principles were developed in the context of Western cultures, particularly the United States (..) The relevance of these principles may vary across cultures, and across sub-cultures within the United States” (Guiding principles, Internet version, p.4).

Our previous table shows how tricky and potentially controversial¹⁰ the “general/public welfare” issue seems in that respect, while at the same time it is probably one of the key issues concerning the relationship between evaluation, ethics and politics.

There have been many ways of formulating aspects and dimensions of the evaluators’ responsibilities with regard to that public concern. This may be addressed for instance in literature as the enlightenment function of evaluation (Weiss, 1993 [1973]), the “societal learning” dimension (Toulemonde and Rieper, 1997), the “cognitive input to society” (Conseil scientifique, 1996) or under the AEA “public welfare” category. Some could even dismiss the question as being to “woolly” and not pragmatic enough to be addressed in terms of standards. We would subscribe anyway to Weiss’s contention that evaluation always take a “political stance” (1993[1973]).

Finally, the formulation of standards should take into account *the diversity of their numerous potential uses*. We think that these are dependent on national variables including the legal framework, the existence of organized interests within the evaluation milieu, but also the nature of the evaluation market supply and offer. In the French case, the following might apply specifically:

- Ethical regulating of the evaluation market (protection of professional evaluators independence and of the general interest);
- Ethical regulating of the political use of the evaluation label in the public debate;
- Strengthening the evaluation professional milieu;
- Clarifying “good evaluation practice” requisites (what is a good evaluation?);
- Fixing priorities for improving methods.

¹⁰ See also the discussion in Patton’s (1997) previously quoted chapter 14.

II - The SFE working group approach

On the basis of these preliminary reflections and exchanges within the working group, SFE's second general assembly (June 2000) was presented with the following analysis.

It appeared to the group that the *procedural quality* of evaluation should be more clearly distinguished both from product quality and personal ethics. In a French and European context where evaluation often bears a more institutional aspect, this issue appeared of particular importance. Evaluation there is frequently implemented through specific institutional settings (including steering groups and sometimes methodological regulation devices) contrasting with a northern American situation, where methodological choices and evaluation credibility rather rely exclusively upon professional evaluators.

For the moment at least, rather than further elaborating on distinctions between standards and guidelines, and ethics and quality, the group thought suitable to adopt a typology of four categories of norms according to their destination:

- 1) Personal ethics of evaluators and other actors of the evaluation process;
- 2) The quality of evaluation processes (relationships between evaluators, clients and stakeholders, evaluation objectives and terms of reference delineation, etc.) from a methodological, organisational and juridical point of view;
- 3) The quality of evaluation products (reports and other evaluation outputs) from a cognitive and scientific point of view (validity, scope of the results...) and from a formal point of view (rigour, impartiality and clarity in the presentation of methods, clear reference to initial questions, readability, etc.);
- 4) The social value of evaluation (its "utility" in the broadest sense, as knowledge production and final product designed to be published).

We will stick to this categorisation for now, because it allows analytical clarification when discussing diverse practical potential uses of standards. We assume (but it could be discussed), that a distinction between the quality of processes and the quality of products is particularly justified in the European - especially French - context, where evaluation is often more institutionalised and proceduralised. The subsequent developments should be taken as a preliminary and transitional attempt to discuss existing sets of standards and guidelines in the view of our categorisation.

Personal ethics and skills

Given their pragmatic approach, the Canadian guidelines appeared suitable to emulate:

CES-SCÉ Guidelines for Ethical Conduct

1. Competence

Evaluators are to be competent in their provision of service.

- 1.1 Evaluators should apply systematic methods of inquiry appropriate to the evaluation.
- 1.2 Evaluators should possess or provide content knowledge appropriate for the evaluation.
- 1.3 Evaluators should continuously strive to improve their methodological and practice skills.

2. Integrity

Evaluators are to act with integrity in their relationships with all stakeholders.

- 2.1 Evaluators should accurately represent their level of skills and knowledge.
- 2.2 Evaluators should declare any conflict of interest to clients before embarking on an evaluation project and at any point where such conflict occurs. This includes conflict of interest on the part of either evaluator or stakeholder.
- 2.3 Evaluators should be sensitive to the cultural and social environment of all stakeholders and conduct themselves in a manner appropriate to this environment.
- 2.4 Evaluators should confer with the client on contractual decisions such as: confidentiality; privacy; communication; and, ownership of findings and reports.

3. Accountability

Evaluators are to be accountable for their performance and their product.

- 3.1 Evaluators should be responsible for the provision of information to clients to facilitate their decision-making concerning the selection of appropriate evaluation strategies and methodologies. Such information should include the limitations of selected methodology.
- 3.2 Evaluators should be responsible for the clear, accurate, and fair, written and/or oral presentation of study findings and limitations, and recommendations.
- 3.3 Evaluators should be responsible in their fiscal decision-making so that expenditures are accounted for and clients receive good value for their dollars.
- 3.4 Evaluators should be responsible for the completion of the evaluation within a reasonable time as agreed to with the clients. Such agreements should acknowledge unprecedented delays resulting from factors beyond the evaluator's control.

Nevertheless some commentaries and questions may be raised as to their relevance in an SFE context:

- a) In the *CES Guidelines*, the definition of integrity does not entail that evaluation should preserve the interests of all beneficiaries and stakeholders. It only requires from evaluators to behave in a manner “appropriate” to the cultural and social environment of evaluation. This approach seems more practical than equally weighing the various interests affected by evaluation, as the *AEA Guiding Principles* seem to imply.
- b) The evaluators’ responsibility only extends to their own contribution to evaluation. This means that an evaluator, who provides a political authority (or an evaluation steering committee) with an evaluation report, should not be deemed accountable for the conclusions drawn and the decisions taken on this basis. He or she should thus have the right, when necessary, to explicitly rule out his/her own responsibility. Do things go that way in practice? What about the risk for an evaluation study to be misused or even cynically used instrumentally? Is it not preferable to assume that evaluators are, to a certain extent, ethically responsible for the dishonest use of their work in the public

debate? This important question remains to be clarified, and case studies could be helpful there.

- c) Do these standards sufficiently take account of the evaluators' heterogeneous institutional positions? Should they apply similarly to private consultants commissioned by a client, to civil servants working within public administration as internal evaluators? Would they be equally relevant for self-evaluation practices and participative or empowerment evaluations? It might be necessary to adapt ethical standards on the basis of a relevant typology of situations.

Process quality and relevance

The purpose of such norms is to characterise good practice with regard to institutional structures, contractual and other formal or informal arrangements between evaluators and their clients, stakeholders, evaluation bodies, steering committees, etc., evaluation mandate formulation, evaluation design. Contrary to the previous category, these standards do not focus on individuals but on organisations and institutions. They envisage evaluation as a social process rather than a professional performance.

Items concerned here are addressed by multiple JCSEE categories. In our perspective it is relevant to index them according to the different steps of the evaluation process:

Preliminary steps

- U1 Stakeholder Identification--Persons involved in or affected by the evaluation should be identified, so that their needs can be addressed
- U3 Information Scope and Selection--Information collected should be broadly selected to address pertinent questions about the program and be responsive to the needs and interests of clients and other specified stakeholders
- U4 Values Identification--The perspectives, procedures, and rationale used to interpret the findings should be carefully described, so that the bases for value judgements are clear.
- P2 Formal Agreements--Obligations of the formal parties to an evaluation (what is to be done, how, by whom, when) should be agreed to in writing, so that these parties are obligated to adhere to all conditions of the agreement or formally to renegotiate it.
- P7 Conflict of Interest--Conflict of interest should be dealt with openly and honestly, so that it does not compromise the evaluation processes and results.

Final steps

- U6 Report Timeliness and Dissemination--Significant interim findings and evaluation reports should be disseminated to intended users, so that they can be used in a timely fashion.
- P6 Disclosure of Findings--The formal parties to an evaluation should ensure that the full set of evaluation findings along with pertinent limitations are made accessible to the persons affected by the evaluation, and any others with expressed legal rights to receive the results.
- A12 Meta-evaluation--The evaluation itself should be formatively and summatively evaluated against these and other pertinent standards, so that its conduct is appropriately guided and, on completion, stakeholders can closely examine its strengths and weaknesses.

Standard U7 applies to the whole process:

U7 Evaluation Impact--Evaluations should be planned, conducted, and reported in ways that encourage follow-through by stakeholders, so that the likelihood that the evaluation will be used is increased

All these standards potentially apply not only to evaluators but to all actors of an evaluation considered as a social process. Hence, the already mentioned diversity of situations should again be addressed. These criteria probably ought to be adapted to different evaluation goals and institutional contexts. On the other hand all evaluations, even internal ones, should meet minimal formalisation requirements (a document should, at least, explicit the objectives, questions to be answered and anticipated use of results). In a sense, all evaluations should come within a contractual framework.

Product quality

This category approximately fits in with the JCSEE “Accuracy” category¹¹:

JCSEE Accuracy standards
<p>The accuracy standards are intended to ensure that an evaluation will reveal and convey technically adequate information about the features that determine worth or merit of the program being evaluated.</p> <ul style="list-style-type: none"> ▪ A1 Program Documentation--The program being evaluated should be described and documented clearly and accurately, so that the program is clearly identified. ▪ A2 Context Analysis--The context in which the program exists should be examined in enough detail, so that its likely influences on the program can be identified. ▪ A3 Described Purposes and Procedures--The purposes and procedures of the evaluation should be monitored and described in enough detail, so that they can be identified and assessed. ▪ A4 Defensible Information Sources--The sources of information used in a program evaluation should be described in enough detail, so that the adequacy of the information can be assessed. ▪ A5 Valid Information--The information gathering procedures should be chosen or developed and then implemented so that they will assure that the interpretation arrived at is valid for the intended use. ▪ A6 Reliable Information--The information gathering procedures should be chosen or developed and then implemented so that they will assure that the information obtained is sufficiently reliable for the intended use. ▪ A7 Systematic Information--The information collected, processed, and reported in an evaluation should be systematically reviewed and any errors found should be corrected. ▪ A8 Analysis of Quantitative Information--Quantitative information in an evaluation should be appropriately and systematically analyzed so that evaluation questions are effectively answered. ▪ A9 Analysis of Qualitative Information--Qualitative information in an evaluation should be appropriately and systematically analyzed so that evaluation questions are effectively answered. ▪ A10 Justified Conclusions--The conclusions reached in an evaluation should be explicitly justified, so that stakeholders can assess them. ▪ A11 Impartial Reporting--Reporting procedures should guard against distortion caused by personal feelings and biases of any party to the evaluation, so that evaluation reports fairly reflect the evaluation findings.

To these can be added one of the “Utility” standards:

<ul style="list-style-type: none"> ▪ U5 Report Clarity--Evaluation reports should clearly describe the program being evaluated, including its context, and the purposes, procedures, and findings of the evaluation, so that essential information is provided and easily understood

And one of the “Propriety” standards:

<ul style="list-style-type: none"> ▪ P5 Complete and Fair Assessment--The evaluation should be complete and fair in its examination and recording of strengths and weaknesses of the program being evaluated, so that strengths can be built upon and problem areas addressed
--

¹¹ Except for A12 (meta-evaluation), related to process quality.

This set of criteria approximately encompasses the same content as the three criteria (“Reliability”, “Impartiality” and “Transparency”) established by the former French Scientific Council for Evaluation (Conseil scientifique, 1996, p. 46).

Conseil scientifique de l'évaluation criteria

Reliability: The evaluation must be trustworthy. This implies accuracy of the information collected and scientific merit for causal deductions upon which the evaluation judgement is based. Particular attention ought to be paid to the bias data collection and processing techniques may introduce into the drafting of conclusions.

Impartiality: The evaluations' conclusions should not be influenced by the personal preferences or institutional positions of those in charge of the evaluation (research supervisors or evaluation body members) or, at least, that any such preferences have been sufficiently explained or examined for it to be supposed that another evaluation responding to the same question and using the same methods would reach the same conclusions. At stake here are the seriousness and honesty of the work done to qualify (should such and such development, for example, be described as «fast», «normal» or «slow») and interpret data by which observation proceeds to judgement.

Transparency: This standard reflects the idea that an evaluation, besides fully and carefully describing the methods employed, should outline its own «instructions for use» and its limitations: position with regard to other possible evaluations of the same subject, résumé of questions left unanswered or incompletely answered, list of possible objections, etc. This attempt at clear-sightedness and self-criticism is necessary to the extent that evaluations are seldom flawless, they leave many questions unanswered and their findings are not always penetrating and unquestionable.

To the extent that they principally apply to evaluation reports, these criteria may be usefully completed by usual recommendations for writing evaluations reports, about the structure, content and writing style of evaluation reports. For example those mentioned in an European Commission guide:

Guide for evaluating EU expenditure (1997)

“A potential reader of an evaluation report must be able to understand:

- the purpose of the evaluation;
- exactly what was evaluated;
- how the evaluation was designed and conducted;
- what evidence was found;
- what conclusions were drawn; and
- what recommendations, if any, were made”.

The same guide stresses the need of “well written” executive summaries, because “it is likely that only a small proportion of the target audience will read the full report” (European Commission, 199x, p. 79). The guide also gives a list of problems, which may alter the clarity of an evaluation report:

- “failing to describe the programme being evaluated in sufficient detail (i.e. assuming that everyone who reads the evaluation report will be sufficiently acquainted with the programme and its rationale);
- failing to describe the methods used in the evaluation for the collection and analysis of data, to justify the choice of methods used or to indicate the strengths and weaknesses of the selected methods;
- using information without giving the source;
- arriving at findings which are not based firmly on evidence;
- reaching conclusions which are not explicitly justified (i.e. not systematically supported by findings), so that an independent reader cannot assess their validity; and
- making recommendations which are not adequately derived from conclusions”.

Very similar recommendations were made by the Conseil scientifique (1996, p. 40). It must be stressed that contrary to others, norms concerning the quality of evaluation reports are fairly consensual and raise no other problem than practical ones (finding a compromise between accuracy and readability).

The social utility and value of evaluation

To discuss this area of norms, the Conseil scientifique (1996) “Utility / Relevance” criteria seem particularly fit.

Utility/Relevance (Conseil scientifique de l’évaluation, 1996)

“The evaluation should produce understandable and useful information not only for policymakers but for all the public policy protagonists. For this to be the case, the evaluation report must respond directly and intelligibly to the questions posed in the initial plan. The standard may also include observance of deadlines. Lastly, the value of the *cognitive contribution* to society is appraised. Ideas suggested by the evaluation may in a general way support public judgement formation on the policy evaluated and add to its information on social issues directly connected with the policy. They can also help research and political thinking on the problems surrounding the policy to advance”.

Two of our categories (product and process quality) overlap with the “social utility” perspective. The same may be said of the seven JCSEE “Utility” standards:

JCSEE Utility standards

The utility standards are intended to ensure that an evaluation will serve the information needs of intended users.

- U1 Stakeholder Identification--Persons involved in or affected by the evaluation should be identified, so that their needs can be addressed.
- U2 Evaluator Credibility--The persons conducting the evaluation should be both trustworthy and competent to perform the evaluation, so that the evaluation findings achieve maximum credibility and acceptance.
- U3 Information Scope and Selection--Information collected should be broadly selected to address pertinent questions about the program and be responsive to the needs and interests of clients and other specified stakeholders.
- U4 Values Identification--The perspectives, procedures, and rationale used to interpret the findings should be carefully described, so that the bases for value judgements are clear.
- U5 Report Clarity--Evaluation reports should clearly describe the program being evaluated, including its context, and the purposes, procedures, and findings of the evaluation, so that essential information is provided and easily understood.
- U6 Report Timeliness and Dissemination--Significant interim findings and evaluation reports should be disseminated to intended users, so that they can be used in a timely fashion.
- U7 Evaluation Impact--Evaluations should be planned, conducted, and reported in ways that encourage follow-through by stakeholders, so that the likelihood that the evaluation will be used is increased.

Nevertheless these JCSEE standards actually address the quality of products and processes rather than the social value and utility in our sense. Besides, the Joint committee “Feasibility” standards could also be related to the social value of evaluation:

JCSEE Feasibility standards

The feasibility standards are intended to ensure that an evaluation will be realistic, prudent, diplomatic, and frugal :

- F1 Practical Procedures-The evaluation procedures should be practical, to keep disruption to a minimum while needed information is obtained.
- F2 Political Viability--The evaluation should be planned and conducted with anticipation of the different positions of various interest groups, so that their cooperation may be obtained, and so that possible attempts by any of these groups to curtail evaluation operations or to bias or misapply the results can be averted or counteracted.
- F3 Cost Effectiveness-The evaluation should be efficient and produce information of sufficient value, so that the resources expended can be justified.

In our perspective, social value and utility should rather be envisaged as a “meta-standard”, aiming at supporting *ex-post* evaluation of evaluations rather than at characterising good practice in an analytic way. It should heed for:

a) as plus :

- relevant answers to initial questions
- practical, credible and general¹² conclusions
- decisional and managerial outputs
- formative outputs
- cognitive outputs
- methodological learning outputs

b) as minus :

- overall financial cost of evaluation
- organisational nuisance it has been the cause of

Conclusion: standards, for which use?

On the basis of the above considerations, the working group is now in a position to proceed to writing some form of “charter” to be discussed within and adopted by the SFE. As has already been alluded to, given the implicit reference of the most influential sets of norms to American and Canadian political culture and institutional contexts, this document ought to at least partially reformulate the various standards to adapt to the French context and organize them along the framework here discussed.

As a provisional conclusion reached by the working group, the hypothesis of adopting a charter on standards and ethics should also be specified in terms of practical use and scenarios. Obviously the process of adopting and then implementing such norms will be gradual. Steps have to be thought of.

¹² Refer to the “Potential for generalisation” criteria established by the French Scientific Council “It needs to be known whether the conclusions reached by observing public action effects in certain circumstances over a given period, can apply generally to other situations and contexts, or even to other policies. In most cases, an evaluation’s findings are used to inform decisions applying to action and contexts that differ in many respects from the ones covered at first”.

The first of these, after the charter is adopted, would be to invite evaluators to subscribe to the charter and to voluntarily abide by its rules. They thus would refer to the charter when answering tenders or proposing their provision of services.

In a subsequent step, this charter could be acknowledged as a valid document when litigation would arise between commissioners and evaluators (for example, in case of unjustified demand of modifying the conclusions of an evaluation report). This step of course is more difficult to achieve.

Notwithstanding the assumption of the charter gradually taking on more juridical weight, it could also be used in the public debate to challenge the usage of evaluation as a label. In this debate, the term is sometimes abusively used to cover undue doctoring of a study's results and to strategically instrument legitimisation of decisions by politicians or administrative executives.

The charter could also be used as a reference by official bodies in charge of validating the quality of evaluations (National Evaluation Council, regional scientific councils). Evaluation steering committees as well as other evaluation bodies could be invited to collectively adopt the charter.

Finally, meta-evaluations could be implemented on the basis of this charter, especially referring to the "social value" criteria.

References

- Barbier J.C., 1999, "Intergovernmental Evaluation: Balancing Stakeholders' Expectations with Enlightenment Objectives?", *Evaluation*, vol 5, 4, October, p. 373 –386.
- Conseil Scientifique de l' Évaluation, 1996, *Petit Guide de l'Évaluation*, La Documentation Française.
- European Commission, 1997, *Evaluating EU Expenditure Programmes*, Brussels.
- Leca J., 1997, "L'évaluation comme intervention, sciences sociales et pratiques administratives et politiques", communication au colloque "L'évaluation des politiques publiques", GAPP-CNRS-ENS, Cachan, 5-6 février, 19 p.
- Patton M.Q., 1997, *Utilization Focused Evaluation*, Sage, Thousand Oaks.
- Toulemonde J. and Rieper O., 1997, eds, *Politics and Practices of Intergovernmental Evaluation*, Transaction Publishers, New Brunswick, New Jersey.
- Weiss Carol H., 1993 [1973], "Where Politics and Evaluation Research Meet", *Evaluation Practice*, Vol 14, n°1, pp. 93-106.
- Weiss Carol H., 1998, *Evaluation*, 2nd ed., Prentice Hall, Upper Saddle River, New Jersey.