

# The Role of Evaluation in Democracy: Can it be Strengthened by Evaluation Standards? A European Perspective<sup>\*</sup>

Wolfgang Beywl  
*Universität Bern*

Evaluation should make a contribution to the further development of politics and society. Evaluation is itself a productive factor in democratic societies and is obliged to function according to its goals and rules.

The general theme around this paper is to discuss possibilities to enhance the contribution that evaluation can make to foster democratic governance. I will address a very small aspect of this big issue. My perspective will be on evaluation as a profession and a community which is on the way to establishing its position in modern democracies, supported by making explicit its own standards for good evaluation practice.

## The Position of Evaluation in Democracy

There are many factors influencing the position of evaluation in the democratic process; some of them are:

- The legal and constitutional bases for evaluation (for example article 170, included in the Swiss constitution in 1999, introduces evaluation as an obligatory task of the federal assembly; see Bussmann 2006)
- The responsiveness of government, administration, political parties to empirically based conclusions and recommendations (competition versus consensus oriented democracy...)
- The degree in which outcome orientation and programming is a natural part of public administration procedures and is called in from service providers who are publicly financed

---

<sup>\*</sup> This article is based on the keynote speech given at 6<sup>th</sup> biennial conference “Governance, Democracy and Evaluation,” 30 September 2004 - 02 October 2004, Berlin, Germany. [<http://www.europeanevaluation.org>]

- The available units in the governmental system that execute evaluation institutionally (Auditor-Generals offices, parliamentary administration...)
- The anchorage of a public discourse, which can be different in democracies with plebiscite elements and high decentralization vs. democracies strongly centralised and/or political parties with great influence on government
- The accessibility of information in the public system; both the accessibility of data and the accessibility of evaluation reports, especially if these are financed publicly (freedom of information/ publics' right to know)
- The supply of evaluation training and studies, particularly in higher education institutions, or relevant courses in political science and public administration studies, etc.
- Dissemination of evaluations in everyday areas of life like school, social services, labour market

All these factors are either not at all or merely in the long run influenceable by the actions of the evaluation community.

A community of evaluators and other experts interested in evaluation, such as the European Evaluation Society (EES) or the national evaluation societies in Europe, has restricted resources and possibilities of influencing these factors via information, exchange of experience, and training, professional standards. Today I want to stress particularly quality guidelines, principles or standards as a framework for the professional quality of evaluation and a basis to communicate the mission of evaluation to the actors of the political system and to the society in general.

To think about evaluation standards is also triggered by risks that evaluation is exposed to. These risks, in the long run, threaten its position in support of democratic policy-making.

Risks for evaluation in this important role become obvious as following:

- Public commissioners: They engage internal units or external companies to do evaluation. These carry out the evaluation in a way in which commissioners themselves, the media, or the political opposition discover quite obvious shortcomings.
- Evaluators: As an evaluator you hand in a tender which satisfies heavy quality demands. The public commissioner decides in favour of the cheapest supplier. Professional requirements on evaluations are known neither to the commissioner nor to the supplier.
- Stakeholders/citizens: As persons affected by evaluation results they doubt the precision and the fairness as soon as conclusions and recommendations have been shared. They don't find any basis on which they can check the quality of the evaluations carried out.
- As evaluation community: Evaluation is in competition to other professions: for example, auditing, controlling, or quality management. These have clear international or European

and widely published quality criteria for the services they deliver. They advertise with these quality codes and thus try to create confidence in the effectiveness of their approaches.

These are some factors which endanger the credibility and the position of evaluation.

What would be simpler in this situation than the following solution? The evaluation community ratifies a clear set of rules which are highly obligatory if individuals or organisations want to belong to this community. It is a mandatory base for training and academic graduation in evaluation. Even courts can refer to these written rules in disputes; accreditation/certification of evaluators or evaluation institutes and also systematic meta-evaluation or evaluation audits can build on this framework.

The outlined solution is based on the assumption that evaluation is a profession like any other: A validated specialized knowledge is available which must be proven again and again; there are obligatory guidelines; certifications which have to be paid for, courts of arbitration, and so on.

Of course, some widely held beliefs about evaluation speak against such a solution: There are different perspectives on what constitutes high-quality evaluation. Respect for particularities and differences are part of what evaluation has come to signify. Europe contains diverse societies, each with its unique political traditions and cultures, understandings of democracy, and ideas about the role of evaluation in a democracy.

We are facing a dilemma: The evaluation community in an expanding unified Europe urgently needs a strong identity, a shared and clearly communicated self-image. As evaluation experts the members of the community are aware of the difficulties to arrive at internationally agreed upon joint principles or standards that adequately reflect multicultural and pluralist political contexts.

Is this an unsolvable dilemma? Or is there a way out to escape these contradictory demands?

I would like to share with you now some thoughts on these questions.

## Evaluation Standards in Europe: A Short Overview

I would like to begin with a short remark on terminology: It has often been argued that it is arbitrary whether to use the term guideline, principle, or standard to address the norms or rules we are talking about. This position is also expressed by the denomination of the EES working group “on evaluation standards, guidelines and principles.” I will follow the example of the working group, which selected the term “standard” provisionally as the working term, including any relevant document that has another title.

In many countries of Europe by now there exist guidelines or standards for evaluation. I list some of them briefly and characterize them with a few words. There are numerous specialists present at this biennial EES conference that have intensively discussed and analyzed these different standard sets. You are invited to study the contributions of these experts and to use them fruitfully in the ensuing discussions (i.e. Widmer, 2004a, 2004b).

- The Swiss Evaluation Society (founded in 1996) has approved the SEVAL Evaluation Standards in its general assembly in 2001.
- The German Evaluation Society (founded in 1997) ratified its DeGEval-Standards also in 2001.
- The French Evaluation Society (SFE, established in 1999) presented a Charter of evaluation guiding principles for public policies and programmes in 2003.
- The UK Evaluation Society (which was the first European Society, founded in 1994) adopted “The Guidelines for Good Practice in Evaluation” also in 2003.

This list is not complete. It concentrates on standard sets of European evaluation societies which have all been established in a process characterized by strong membership participation, including hearings, expert groups, and discussions at annual conferences and so on. Other examples include the Italian Evaluation Association that developed a document parallel to its founding process in 1997 and there are also Standards laid out by the Finnish Evaluation Society. The Spanish evaluation community is also thinking about standards or guidelines (Bustelo, 2004).

There are other important sources explaining the requirements for good evaluations:

- The EC Evaluation Standards and Good Practice Guidelines are an important element of an evaluation guide, published in by the European Commission - DG Budget (2003).
- The MEANS collection (European Commission - DG Regional Policy 1999) included standards in eight dimensions. In this tradition 44 “Golden Rules” were published about good evaluation practice of socio-economic development programs within the four chapters of “The Guide”, internet-published by the European Commission – DG Regional Policy (2003). This is an internet based comprehensive evaluation handbook sponsored by the DG Regional Policy.

There are many other guidelines and standards, some of them explicitly formulating requirements for evaluations in distinct evaluation fields. On the European level there are for example the Guidelines for the Evaluation of Drug Prevention, published by the European Monitoring Centre for Drugs and Drug Addiction (EMCDDA, 1998). In Austria a group of experts developed Standards for technology and research evaluation, accompanied with checklists, guidelines for terms of reference and many other tools (FTEVAL, 2003). In Switzerland for instance the Evaluation Management & Resource Centre within the Swiss Federal Office of Public Health has published the Guidelines for Health Programme & Project Evaluation Planning (CCE, 1997). Also national Auditors-Generals offices continually work on norms and guidelines how to integrate effectively the evaluation function.

Whereas most sets of standards claim applicability to evaluands in *all* sectors (private, state and third sector) some focus on public activities, as the SFE Charter and also the EC Standards and

Good Practice Guidelines (Widmer, 2004a). An important focus is on *programmes*<sup>1</sup> as evaluation objects but also on experimental legislation and other interventions or actions of public or private bodies. SEVAL- and DeGEval-Standards also apply to organisations/institutions and policy systems (as for example a state-wide higher education system). They apply to a wide variety of evaluation fields, including education, health, social services, consumer protection, research, criminal justice, development cooperation, human rights, gender mainstreaming or social inclusion strategies and so on and so on. And they apply to (programs of) governmental bodies, Parafisci, market enterprises and to various kinds of third sector collectives such as social welfare organisations, foundations or any non-governmental organisations.

Why this long list? I want to underline that there is much common ground for standards in Europe, standards that have a kind of an umbrella function, covering many fields of evaluation, types of evaluands, within and outside the control of public authorities. So most of them are not specifically established with the explicit aim to guide evaluations within democratically controlled systems, but they are well suited to fulfil this task and they comprise many elements which specifically adapt evaluations to the traditions, needs and procedures of democratic societies.

To draw one first conclusion: Within many of the European countries there is obviously a great need for evaluation standards and guidelines. Some of the national evaluation societies have already ratified—often preceded by a broadly arranged participatory process—bodies of norms or standards. These might have an effect on evaluation culture, primarily in the public sector. Public authorities themselves—on the European as well as on the national (maybe at the regional) level—also develop and optimize rules and guidelines for evaluation, which are sometimes specific to distinct evaluation fields.

As a response the EES-board has recently started to develop a policy on standards, guidelines, codes and principles in evaluation. On the EES Web pages a draft paper by Thomas Widmer is published which includes three essentials regarding the cooperation between EES and national evaluation societies concerning standards, guidelines and principles:

- Promotion: that is, information on standard setting procedures employed by various national evaluation societies in Europe and other organisations as well as public authorities
- Exchange: instead of development of ‘their own’ EES standards facilitation of mutual exchange
- Plurality: No preference for any kind of standards, guidelines and so on but openness to ongoing developments.

At the EES conference, there were some sessions organised by the recently established standing EES working group on “standards.” As I read in the abstracts booklet there are many informed

---

<sup>1</sup> Programme as a generic term comprising for example education or training curricula, European structural funds measures, prevention interventions, media campaigns, and so on.

specialists who do presentations on this theme, so that you can get in-depth information on various important aspect of this ongoing debate.

Obviously the EES has chosen—as in many other areas and conforming to its statutes and philosophy—to play a subsidiary role in the question of standards for evaluation in Europe.

To me there is a tension between great enthusiasm for many different sets of standard, guidelines, and principles, expressing cultural and policy field diversity, on the one hand, and, on the other hand, there is the threat of segmentation of a quite small professional community in Europe, as a consequence of a swarm of different papers (specialized on evaluation fields, on kind of evaluation...), with only slightly different orientation but varying terminology. This results in the need to build elaborate translation glossaries and transformation grids to avoid provoking misunderstandings constantly.

In the remainder of this paper I want to discuss some arguments against standards, list some of their potential benefits, reframe some of the objections from a utilisation perspective, and highlight the question of values consideration in evaluation and its standards.

## Arguments Against Evaluation Standards

There are many objections to standards for evaluation. For the most part they express fears that should be taken seriously. However, I see a danger that these objections are also sometimes expressed for reasons of political correctness. Of course it is an important issue to be aware of cultural diversity, different evaluation traditions, questions of gender and ethnicity when formulating guidelines for good evaluations. Of course we will never find a perfect standard set that fits all possible evaluation tasks and socio-political contexts equally. I believe it is important to discuss these issues and find the best solution, written down and communicated to the evaluation community and to the public.

Objections *against* standards are based on distinct interests and values; it is the same for arguments in favour of standards. These defences and objections apply to the central value claims (as Ernest House would call them) of the evaluation profession, art or science (alternative names for what we do, which also include value claims) itself. It would lead too far here to address all of them. I only want to list some of them exemplars:

- a) The same set of standards does not fit into substantially different political cultures. Strongly centralised democracies need other evaluations than decentralised ones; the same is true for consensus-oriented democracies on the one hand, conflict-oriented democracies on the other hand; democracies with full time professional members of parliament have another evaluation utilisation mechanism than others with mostly volunteer politicians; there are small political systems with less than 1 million inhabitants and others with more than 30 million.
- b) Standards do not handle value questions adequately, which are the essence of evaluation. Sometimes they ask for *transparency* of values, interests and resulting evaluation criteria. But is this enough? We know that distinct interests and values shape evaluation contracts,

evaluation questions, interpretations and the utilisation of evaluation findings. Standards often give no clear orientation as to which is the sound way to identify and justify values, which way conforms to ethical or to democratic principles. Value-ignorant standards would be seen as a surrender of evaluation itself.

- c) Existing standards often have a strong bias concerning distinct kinds or settings of evaluation that is implicit and not explained openly: Some of the best known standard sets are written from the (idealised) perspective of an external independent evaluator removing the preconditions of internal and self evaluations. Others privilege evaluations done within democratically controlled settings as in the public sector, neglecting the requirements of evaluation done for the private sector.
- d) Maybe standards can improve the work executed by evaluators, at least a little bit, but they merely pretend that evaluation has a big influence on the decision makers and program directors. Standards make no sense if commissioners say no to broad stakeholder identification, to transparency of values, or to disclosure of findings—which is not a single standing event, at least in my own evaluation practise. Standards—a toothless paper tiger?
- e) There is no standard set thinkable that is suitable for all audiences. Some evaluators need clear instructions, maybe in technical language, how to execute evaluations, others dislike. So for them maybe adequate standards should be organised along the phases of evaluation starting with tender/offer and ending with follow up/meta-evaluation. Clients on the other hand urgently need regulations concerning contracts, as well as costs and benefits of evaluation. Politicians mainly need information about credibility and impartiality of evaluations. The general public wants a short text that is easy to understand so that an illustrative picture of good evaluation arises.
- f) Standard sets often include several single standards that contradict each other: For example political viability/diplomatic conduct as precondition for feasibility of evaluations/the survival of the evaluator as a contractor on the one hand, independence/isolation from vested interests as precondition of a fair and complete assessment of a credible/honest evaluator on the other hand. The argument is: If not all standards are applicable, evaluation practice becomes ethically and professionally puzzled—such kinds of standards would encourage an arbitrary evaluation practice and would be counterproductive to its own mission.
- g) Finally, there is the claim that standards impede innovation, because they fix in a prescriptive way what is good practice and what is bad practice/not allowed. Rigorous guidelines could thus hinder progress in evaluation theory or research methods. Members of the evaluation community who strictly adhere to the guidelines would tend to overlook requirements for new kinds of evaluation caused by social or political change.

## Benefits and Uses of Evaluation Standards

There are many benefits promised by supporters of evaluation standards. I want to mention some of them: Standards...

- Foster the dialogue between the evaluation community and the general public, politicians and administrators about the potential role of evaluation in democracy and society;
- Support the creation of an evaluation culture in public administrations and private organisations as well;
- Strengthen the identity of evaluation as a profession, art and/or science;
- Provide evidence on the progress of evaluation theory, its key terms and concepts and at the same time promote theoretical and methodological development;
- By incorporating key definitions of evaluation language, such as “values,” “interests,” “stakeholder,” “audience,” “target group,” “criteria,” “indicators”...make communication about evaluation easier and more precise;
- Provide a framework to shape the relations between clients/commissioners, directors/program managers and evaluation teams; and clarify their roles;
- Serve as textbooks for evaluation training; give novices clear orientation on their way to mastery; allow evaluation experts to assess their degree of professionalism and to identify developmental needs; and
- Function as a practical guide on how to do an evaluation, step-by step within its main phases (defining evaluation purpose and questions, collecting and valuing empirical information, communicating and reporting findings).

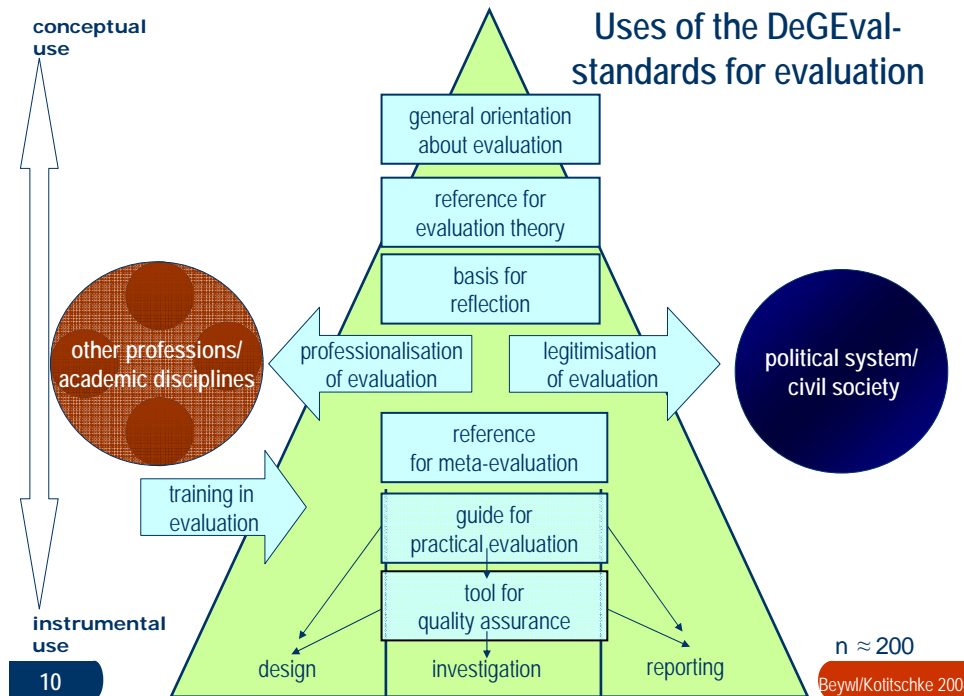
This list sounds promising. The problem is that we have no empirical evidence about the effects of standards, their measured utilisation, their counter-effects and their cost-benefit relationship.

I cannot fill this gap—as an unsatisfactory substitute I would like to present the first results of an online survey concerning the DeGEval-Standards (see DeGEval 2004). These findings lead me to the idea that different intended uses of evaluation standards need to be adequately reflected in a standards text, in addition to providing the standards in specific, audience-tailored formats, for example for politicians or the general public as the primary intended users.

Some words about the Online-survey: The questionnaire was put online in spring of 2004. Two-hundred fifty-seven people participated. As a result of the dissemination paths of the invitation to participate—mainly by E-Mail, also a snow-ball procedure—about 55% of the respondents are DeGEval-members; that is about 30% of the total membership. The main purpose of the inquiry was to get feedback from evaluation specialists about the timing and scope of a revision process of the DeGEval-Standards, but also to gain some evidence about kinds of their utilisation. The results will be published at the end of 2004.



Question 4 of the survey asks the following: “In which regard are the standards useful for your work? Would you please outline briefly for which purposes you use the standards.” Most of the answers reflected in the following analysis belong to question 4. Some belong to other related open ended questions (asking for the preferred way to do the revision process or for arguments to choose one of them—confusing, leave out if possible).



The first idea to structure the categories resulting from the content analysis was to bring them in an order from the more general to the more specific. In analogy to Rossi, Freeman, and Lipsey (1999, p. 432)<sup>2</sup> it is possible to differentiate between (more) conceptual and (more) instrumental uses of evaluation standards.

- Conceptual use means the variety of ways in which evaluation standards *indirectly* have an impact on evaluation as a theory and profession and on its recognition by society and politics, stimulating “the thinking about issues in general way.” Standards give an overall orientation about the goals and concepts of evaluation, are expressions of and references for evaluation theory and serve as a basis for reflection for the individual evaluator (or evaluation teams/units/communities) to think about his or her evaluation practice.
- Instrumental use means the *direct* and observable use—here—of evaluation standards: they can serve as criteria for meta-evaluations; they can be the basis for evaluation training curricula and serve as self-instruction material; they can be refined as reference

<sup>2</sup> I use the dichotomy of Peter Rossi and his colleagues for classifying uses of evaluation standards; they use it for classifying uses of evaluation itself.

system for quality assurance—especially concerning collection and analysis of data; and they can serve as a tool for evaluation practice.

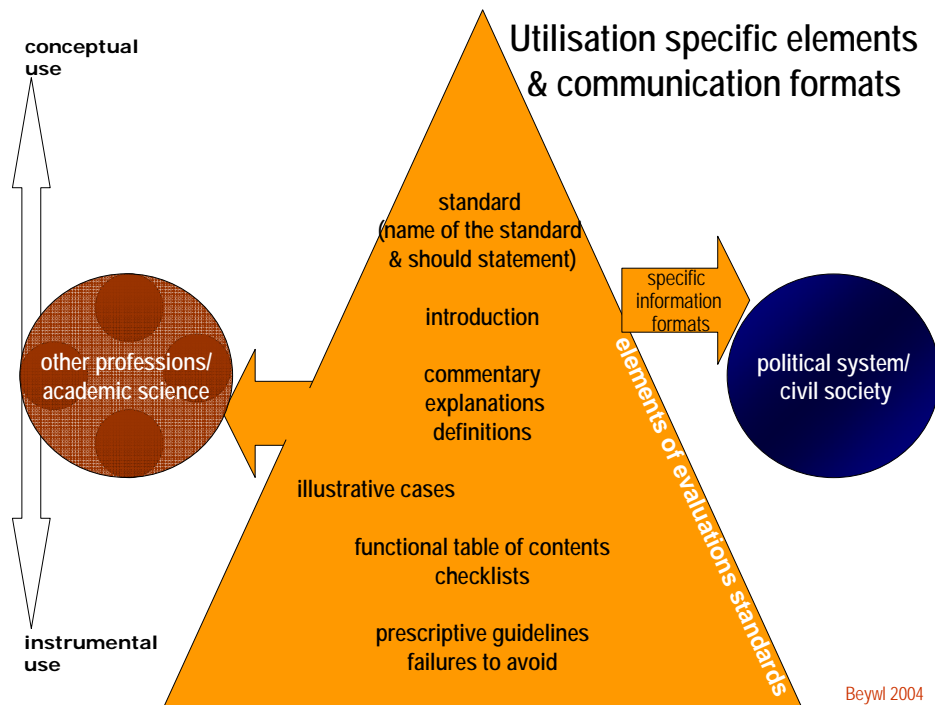
- In between one could position two other uses, often mentioned by the respondents: Standards can enhance the legitimacy of evaluation in relation to the political system or civil society. They can support the willingness of public authorities and politicians to commission evaluations or—one step further—to ascribe evaluation a productive role in the democratic process. Standards can be a weighty point and a visible expression of professionalization of evaluation. They can support evaluation’s positioning vis-à-vis academic research and other professions as for instance accounting or quality management.

Please don’t misunderstand me. What I try to do is a kind of descriptive mapping of the uses of evaluation standards. It is not an attempt to fulfil the need for empirical evidence on the extent of their conceptual and instrumental use. This topic in fact demands substantial research. Maybe you will soon hear about such empirical evidence reported in the EES standards working group.

## Different Standard Elements for Different Uses

This picture of potential uses of evaluation standards makes it possible to untangle some of the arguments against standards though it does not mean to invalidate them. But maybe it is possible to gain new perspectives and find better solutions.

If we see the different uses of evaluation standards—between the two poles ‘conceptual’ and ‘instrumental’—it becomes obvious, that different elements of evaluation standards or guidelines serve different uses (with different intensity):



- The first core elements are the standards themselves. They are usually one to three sentences long and express the desirable attributes of evaluations. These statements are in the centre of the French, the German and the U.K. codes and are called “standards” (CH, D), “principles” (F), or “guidelines” (GB). These texts are highly condensed and communicate a vision of ‘how sound evaluation should be.’ (The standards-statements themselves most clearly feed into the conceptual use of the sets of standards.)
- A second core element is the introduction to these standard texts, providing some short information on the scope, the purpose and the intended uses of the paper.

These two elements give a general orientation about evaluation and include central theoretical, ethical and methodological assumptions and principles. Although short and written in mostly non-technical language they are fully decodable primarily for evaluation specialists, evaluators or informed commissioners/stakeholders. If evaluation wants to communicate its vision and mission to the broader public it needs specific information formats which are—as far as I know—not available within the different evaluation societies.

- A third element is a detailed commentary—about the standards text as a whole and also about each single standard. It addresses the emergence, the application and the future development of the text. The text explains key concepts—such as formative/summative, internal/external—and gives core definitions. This element, which easily comprises 20 or more pages, requires intensive reading and analysis.
- A fourth element are illustrative cases for each of the single standards, examples of ‘good’ and ‘bad’ evaluations which enable the readers to work through the texts and apprehend more complex and specific aspects. Only the Joint Committee (1994) standards include—all in all some 40—illustrative cases. They are often used in evaluation trainings and for many of the respondents of the DeGEval Online Survey such case examples would be the most valuable complement of a new edition.
- The fifth element: practical tools derived from the standards using their structure and logic. The SEVAL and the DeGEval Standards have adopted the “functional table of contents” of the Joint Committee Book, showing how selected standards correspond to each evaluation task (from ‘Deciding Whether to Evaluate’ to “Reporting the Evaluation”). Checklists are also included within these three sets.
- The sixth element is the only one which has a stronger prescriptive character so that it can serve as a possible starting point for quality assurance. (This element thus supports the most clearly instrumental use of the Standards.) The JCSEE give some six to ten prescriptive guidelines and also lists failures to avoid for each of its 30 single standards.<sup>3</sup> Evaluators can use these recommendations and warnings for checking their design and implementation; meta-evaluators find a detailed criteria list for evaluation.

---

<sup>3</sup> Some of the 60 statements of the UKES-Standards could be categorised as more prescriptive guidelines.

There are some clues in the DeGEval-survey that many of the respondents would like a standard book comprising all six elements—but some of them express the concern that this would be an effort overstretching the resources of the German evaluation society.

## Objections Seen from a Utilization Perspective

Some of the arguments against standards mentioned earlier can be reframed by analysing the different elements of evaluation standards and their potential uses.

A first assertion is that none of the above mentioned elements of the existing standard sets in evaluation is comparable to the minimum standards known from technology and industry. Even the “prescriptive guidelines” mentioned as a sixth element are far away from a “minimum standard” as known from the realm of quality assurance or engineering. Minimum standards should not be confused with “maximum standards” as they are widely used in evaluation standard sets. It is crucial to differentiate between these two types of standards:

- A ‘minimum standard’ states, usually in rather technical terms, specific (ideally quantitatively operationalized) minimum requirements, which must be strictly fulfilled so that quality can be ascribed to the object. Minimum standards are strongly prescriptive.
- A ‘maximum standard’ states, usually in everyday language open to interpretation, the envisioned ideal that an object should fulfil in order to be judged as good, high quality etc. Maximum standards give orientation.

Objection g) —standards as an impediment to innovation—is certainly a realistic concern for the case of minimum standards. Texts such as the evaluation principles, guidelines or principles are—in all their elements—clearly based on maximum standards, where innovations are a built-in feature, as for example the UKES Standards express very clearly:

The guidelines are prescriptive only in the sense that they rehearse what those engaged in the practical business of evaluation, from whatever perspective, have found to be both an honourable and effective way of interacting. We believe the guidance will come alive through use in the discussions and negotiations between people involved in evaluations. As such we hope the statements will promote conversation about evaluation in general but also support ways of negotiating some of the critical aspects of the evaluation process from commissioning to dissemination of evaluation findings

So I think that one of the objections can be explained by a misunderstanding about the nature of maximum standards. This makes evident how important it is to define key terms very clearly.

Objection f)—standard sets often include single standards that contradict each other—for at least a part can be reframed using the concept of maximum standards. These are loosely defined statements leaving space for interpretation, negotiation and concretisation depending on the specific purpose of the evaluation in question and the situational context. For example, at least two of the four standards groups of the JCSEE Standards—feasibility and accuracy—in the real

world of evaluation, may in rare cases both be met for 100 percent. Very often compromises will be necessary. There is a certain probability that a ‘perfect’ in the sense of non-contradictory set of standards would indeed impede innovation within our lively profession. But nevertheless: *avoidable* contradictions of standards should be removed so that they give the most possible clear guidance.

Objection e)—there is no standard set which is suitable to *all* audiences—is true and is not true as we can see now. As any communication and any report of a given evaluation have to be precisely adapted to the information needs and capacities of the intended users, so it is the same for evaluation standards. Different elements of a standards textbook are relevant more or less for diverse user groups. If evaluation wants to reach the general public special formats are required additionally.

Objection d)—no obligation of clients/commissioners/public officials to work with evaluation findings—this is a really long-lasting complaint at evaluators’ meetings and in evaluation literature. It should become clear that evaluation standards strive for *conceptual* utilisation by decision makers, public administrators and also politicians. They are a framework within which to negotiate evaluation tenders, re-negotiate evaluation designs and evaluations questions, prepare public presentations of findings, shape communication strategies to disseminate findings, conclusions and may be competing recommendations. I believe maximum standards can strengthen the role of evaluators, give a professional backing for conflict situations and by the way support the process of democratic decision making.

Objection c)—no unique standard set is adequate for different kinds of evaluation—is also to be taken seriously. It is obvious that the perspectives of external evaluators (and to some extent also of commissioners of external evaluations) are well represented within the standard sets. Needs and requirements of internal evaluation, peer-evaluation and self evaluation could be further addressed within the standards or at least its supplementary documents. Maybe internal and self-evaluators are not so well represented on the boards and standards-related standing working groups of national evaluation societies so that it could be necessary to initiate activities in this direction.

There are already examples of how to enlarge the scope of standards to these kinds of evaluation. The UKES guidelines from the very beginning include a section with 17 guidelines for institutional self-evaluation. The DeGEval-Standards will be supplemented by “recommendations to apply the standards on evaluation in self-evaluation.”<sup>4</sup>

I would like to address now objection b)—the value topic in evaluation standards—in more detail and leave argumentation on objection a)—mismatch of standards to distinct democratic systems—to further discussions. But I am sure that reflections about values in evaluation could also inform the discussion about the applicability for diverse democratic systems. Nonetheless

---

<sup>4</sup> There is a draft version of these recommendations now which will be discussed and most likely ratified by the DeGEval general assembly in November 2004.

the latter point needs more research and discussion—that could be a topic to be tackled by the European evaluation community.

## Evaluation Standards and Values in Democracy

The objection that evaluation standards do not adequately handle value questions is a substantial point of argument. As democratic governance has at its core the struggle of diverse interests and values, the representation of values becomes crucial when applying standards to evaluations of public policies.

The four evaluation standard sets mention values in a more or less implicit manner. Both the UKES and the SFE papers touch upon them by demanding that different stakeholder perspectives should be taken into account.<sup>5</sup> The SEVAL and the DeGEval-Standards U5 “Transparency of Values” approach the topic more explicitly: While the SEVAL-Standards—in the tradition of the respective JCSEE-Standard U4—focus on interpretation of findings, the DeGEval-Standard U5 expands the call for transparency of values to all phases of an evaluation:

The perspectives and assumptions of the stakeholders that serve as a basis for the evaluation and the interpretation of the evaluation findings shall be described in a way that clarifies their underlying values.

Within the short commentary given there is a first hint concerning the nature of values:

Interpretation of the collected information and findings in the final phase is one of the most important and critical parts of the evaluation process. Societal values (norms) necessarily play a major role in this. The underlying values shall be as transparent as possible so that interpretation is convincing, comprehensible and assessable.

As an expansion for any evaluation standard set it could be helpful do define the concept of values. This would be in accordance with the function of standards to define generic key terms and to support communication about evaluation concepts.

There are many suggestions in the evaluation literature, especially in the writings of Ernest House who has discussed the topic since the middle 70s. Following Campbell he describes ‘assertions on facts’ (factual claims) as central for the planning and execution of evaluations and supplements these with “value claims.”

Descriptively, values could be defined as belief-based preferences of individuals and collectives to favour distinct settings or solutions over others. Values—normatively seen—express demands put forth to other individuals, to collectives and to the democratic powers, as well as demands to make distinct choices. Thus, the normative function of values is to argue and to enforce that

---

<sup>5</sup> SFE: “L’évaluation prend en compte de façon équilibrée les différents points de vue légitimes qui ont été exprimés sur l’action évaluée.” UKES: “The guidelines provide a reference point from different perspectives for a range of stakeholders involved in the evaluation process.”

specific preferred alternatives are to be selected. Evaluation in the context of democratic governance should make transparent how it handles such value claims.

Ernest House complains that value claims and factual claims are melted and mixed up in evaluation again and again. Therefore, he demands that we develop systematic procedures to collect and analyse claims that include strong value aspects, so that evaluative conclusions are not being biased by underlying implicit value claims. (House/Howe 1999, 313)

Michael Scriven recently underlined his position that value clarification is a definitional element of professional evaluation:

The core of this package consists of the techniques that are involved in the systematic and objective validation of evaluative claims which is the dictionary definition of evaluation with two qualifiers in front of that to narrow it down to professionally competent evaluation (Donaldson & Scriven, 2003, p. 30)

But—is there really a consensus in evaluation—to incorporate values clarification into the evaluation process? Or would we be in danger of privileging—by way of the Trojan Horse “Values Clarification”—some evaluation traditions and to exclude others that are very sceptical with regard to values as the driving force of the evaluation process? For example Rossi and colleagues (Rossi, Freeman, & Lipsey, 1999, p. 422) take the position—being fully aware of the post-modern criticisms:

...that disagreements among researchers on empirical findings are mainly matters of methods or measurement error rather than matters involving different truths...Indeed the message of this book is how to choose the best method for a given research question that is likely to produce the most credible findings.

From my point of view I have no doubt at all that Rossi, Freeman, and Lipsey (1999), Thomas Cook (Shadish, Cook, & Campbell, 2002) or James Heckman (Heckman & Smith, 1996) and many others who advocate methods-driven evaluation, as well as Ernest House, Bob Stake, Michael Scriven and many others who push the values envelope, all have honourable motives to take their stance on how evaluation should handle the value question—a question that accompanies evaluation since its early roots at the beginning of the 20<sup>th</sup> century. This is—as I mentioned in the beginning—a value claim of evaluation itself and it should be handled with great circumspection.

If standards for evaluation are not to be silent at this point—which would offend postmodernist thinkers—or if they take the stance that values are the advance organisers of any evaluation effort—which could enrage another fraction of methodologists, what can be done?

The UKES-Standards that strive for “both an honourable and effective way of interacting” give a very important piece of advice:

There is no evaluation stance for which these guidelines are inappropriate or inapplicable. Many of the statements have at their heart the need to be open and transparent about the expectations and requirements of all the stakeholders

whoever they may be. As such the language used has striven to avoid hidden or tacit assumptions about the efficacy, dominance or normality of any single approach to evaluation.

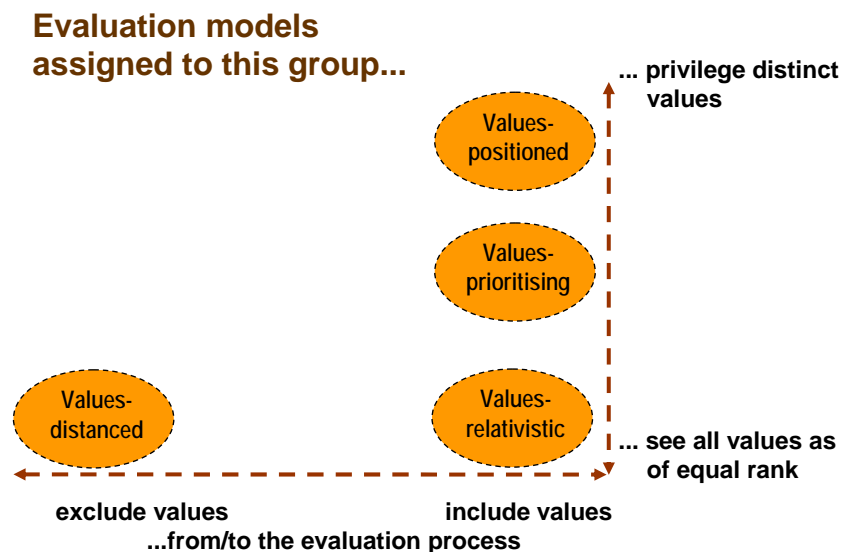
I would like to make a modest proposal which does not provide a solution but which could help clarify the values topic for the evaluation community and for its audiences in politics and society. F so that it becomes transparent how distinct evaluations deal with the values issue so that stakeholders can adequately use findings, conclusions and recommendations.

The following statement or a similar one could be included in an evaluation standards paper:

Evaluators should make explicit how they handle competing interests and values shaping the evaluation design and guiding the presentation of their findings and interpretations, especially whether they include or exclude value clarification into the evaluation process and whether their evaluation models adhere to specific values or rather see all values as of equal rank.

Evaluation models can be ordered along two dimensions: values inclusion versus exclusion and equivalency of all values versus preferential treatment of distinct values.

### Evaluation models categorised as they consider values



The following outline of the four main types is succinct and is no substitute for thorough analysis.<sup>6</sup> Categorisation is guided by the evaluation model's consciousness of values.<sup>7</sup>

<sup>6</sup> An initial systematic portrayal is found in Beywl et al. (2004), dealing primarily with evaluations of poverty avoidance and social inclusion policies and programmes.



Commonly there are overlaps between categories, which result from ambiguities in model descriptions, particularly when the subject of values is merely treated implicitly.

- Values-distanced approaches follow the tradition of thinkers such as Max Weber or Karl Popper and eliminate value judgements from the evaluation process. Theoretical framing of an evaluation and implementation in empirical investigations operate ‘objectively’ according to strict rules; the utilisation of evaluation findings is delegated to the external and public democratic processes which align them with their own competing values.
- Values-positioned approaches explicitly assume that societies are marked by stark power imbalances and social and economic inequality. Evaluations should counterbalance the value hegemony in the political and cultural spheres by strengthening the weak and giving them an audible voice in the political process.
- Values-prioritising models also assume disequilibria in the power distribution of stakeholders, but restrict themselves to making them transparent and accessible to the negotiation of particularly relevant/socially accepted values. For instance, they may demand involvement of all stakeholders in the determination of questions and discussion of findings and may work toward prioritisation and a minimum consensus.
- Values-relativistic models underscore the dominant significance of values in planning, executing and utilising evaluations. They detect value conflicts in all phases and maintain existing tensions without taking sides or making pragmatic compromise. Motivation and social energy in using evaluation findings derive from consciously and publicly stated differences in values and interests among stakeholders.

I will not go into the details of this classification, which owes much to Daniel Stufflebeam (2001) on evaluation models, which takes a distinctly normative position, which I try to avoid. Unquestionably there are many mixtures and overlaps regarding the four categories in the real world of evaluation. The following table gives an idea of how one could assign several well known evaluation models to the four categories—this one does not claim to have mutually exclusive categories either:

---

<sup>7</sup> Assignments are not performed analytically, by maintaining, for instance, that cost-benefit analyses are bound *ipso facto* to the value judgements of shareholders (a stakeholder subgroup) or that goal-free evaluations mainly reflect values that are widespread in society (thus confirming the value hierarchy). Such mutually critical analyses form the nucleus of the ‘paradigmatic debates’ in evaluation methodology (Guba and Lincoln, 1997; Pawson and Tilley, 1997; *Philosophies and types of evaluation research* authored by Eliot Stern in VET).

Table 1  
Evaluation models and how they take into account values

Values consideration	Chapter	Generic name of the model	Synonyms
1. values distanced	2.3.1.1	objectives-focused	effectiveness studies
	2.3.1.2/3	(quasi) experimental design focused	gold standard
	2.3.1.4	cost-benefit- focused	cost-effectiveness
	2.3.1.5	context-mechanism focused	realistic/realist evaluation
	2.3.1.6	program-theory focused	theory-driven evaluation
2. values relativistic	2.3.2.1	issues focused	responsive
	2.3.2.2	Dialog focused	constructivist/fourth generation
3. values prioritising	2.3.3.1	decision focused	accountability oriented
	2.3.3.2	utilisation focused	pragmatic
	2.3.3.3	stakeholder interests focused	deliberative democratic
4. values positioned	2.3.4.1	self organisation focused	empowerment evaluation

My idea is that a high degree of transparency about how distinct evaluations treat the question of values can support credibility, integrity and propriety of evaluation as a profession.

## Outlook

It is my belief that evaluation needs support and backup to strengthen its role in the changing European democracies. Maybe standards are promising to shape a European evaluation identity based on diversity.

I stand up decidedly for evaluation standards that win the support of a wide variety of national societies and organizations, many different democratic thinking traditions, and a wide range of evaluation cultures. The standards should be applicable to all public policy arenas. And they should also apply to private and third sector programmes and organisations as well. Such a document should be supported by evaluation professionals as well as by evaluation commissioners. Some parts of the text should be written in a language that is accessible to journalists so that it can be communicated to the general public.

The existing texts on standards, principles and guidelines of evaluation in Europe have a broad range of common understanding which could be extended.

My vision is a cohesive body of evaluation standards, guidelines or principles developed with active participation of evaluators and evaluation users from a wide range of countries and policy fields, with diverse evaluation cultures and methodologies in mind. A short paper comprising essential statements should be supplemented by explanations, definitions and commentaries, making explicit evaluation theory and evaluative thinking. The book or online data base I have in mind would also include practical tools derived from the standards, such as checklists or working grids. Well written illustrative cases should be included, especially for teaching evaluation and self-instruction of interested parties to open the readers' minds for deeper understanding of evaluation knowledge, methods and performance.

Such a body of text derived from the existing standards, guidelines and principles could be supportive in making evaluation trustworthy and *visible* as an essential part of democratic governance.

## References

- Beywl, W., Speer S., & Kehr, J. (2004). *Wirkungsorientierte Evaluation im Rahmen der Armuts- und Reichtumsberichterstattung. Perspektivstudie*; Bonn; published by the German Ministry of Health and Social Affairs (BMGS). Available at [http://www.univation.org/index.php?class=Calimero\\_Article&id=54](http://www.univation.org/index.php?class=Calimero_Article&id=54)
- Bussmann, W. (2006). *Gesetzesevaluation und experimentelle Rechtsetzung*. In: *ius.fullNo 1/2006* (pp. 42-52).
- Bustelo, M. (2004). The potential role of standards and guidelines in the development of an evaluation culture: The case of Spain. Paper presented at the 6<sup>th</sup> Conference of the European Evaluation Society. Berlin.
- CCE (Centre de Compétences en Evaluation). (1997). *Guidelines for health programme & project evaluation planning*. Available at <http://www.bag.admin.ch/cce/e/index.htm>
- DeGEval (Gesellschaft für Evaluation) (2004). *Revisionsprozess der standards für evaluation*. Available at [http://www.degeval.de/index.php?class=Calimero\\_Webpage&id=9027](http://www.degeval.de/index.php?class=Calimero_Webpage&id=9027)
- Donaldson, S. I., & Scriven, M. (2003). Diverse vision for evaluation in the new millennium: Should we integrate or embrace diversity. In S. I. Donaldson & M. Scriven (Eds.), *Evaluating social programs and problems: Visions for the new millennium* (pp. 3-16). Mahwah, NJ: Lawrence Erlbaum.
- European Commission—DG Budget (2003). *Evaluating EU activities: A practical guide for the commission services*. Brussels: European Commission, 2003. Available at [http://ec.europa.eu/budget/evaluation/pdf/pub\\_eval\\_activities\\_full\\_en.PDF](http://ec.europa.eu/budget/evaluation/pdf/pub_eval_activities_full_en.PDF)
- European Commission—DG Regional Policy (1999). *The MEANS collection. Evaluating socio-economic programmes*. Luxembourg. Off. for Official Publ. of the Europ. Communities. Available at <http://www.eureval.com/publications/means.htm>

- European Commission—DG Regional Policy (2003). *The guide. Evaluation of socio-economic development*. Available at <http://www.evaled.info/>
- EMCDDA (European Monitoring Centre for Drugs and Drug Addiction). (1998). *Guidelines for the evaluation of drug prevention*. Lisbon. Available at <http://www.emcdda.eu.in>
- FTEVAL (Plattform Forschungs- und Technologieevaluierung). (2003). *Standards der evaluierung in der technologie- und forschungspolitik*. Vienna. Available at <http://www.fteval.at/standards/>
- Heckman, J. J., & Smith J. A. (1996). Experimental and nonexperimental evaluation. In G. Schmid, J. O'Reilly, & K. Schömann (Eds.), *International handbook of labour market policy and evaluation* (pp. 37-88). Cheltenham. Edward Elgar.
- House, E. R., & Howe, K. R. (1999). *Values in evaluation and social research*. Thousand Oaks, CA: Sage.
- Joint Committee on Standards for Educational Evaluation (1994). *The program evaluation standards* (2<sup>nd</sup> ed.) Thousand Oaks, Calif.: Sage.
- Rossi, P., Freeman, H. & Lipsey, Mark (1999): *Evaluation: A systematic approach* (6<sup>th</sup> ed.). Thousand Oaks, CA: Sage.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.
- Stufflebeam, D. L. (2001). *New directions for evaluation: Evaluation models* (Vol. 89). San Francisco: Jossey-Bass.
- Widmer, T. (2004a). *EES policy on standards, guidelines, codes and principles in evaluation* (Draft proposal). Available at [http://www.europeanevaluation.org/docs/standardspolicy\\_v2\\_18022004.pdf](http://www.europeanevaluation.org/docs/standardspolicy_v2_18022004.pdf)
- Widmer, T. (2004b). The development and status of evaluation standards in Western Europe. *New directions for evaluation* (Vol. 104, pp. 31-42). San Francisco, CA: Jossey-Bass.