



Setting Standards and Providing Guidelines

The Means Toward What End?

MARLÈNE M. LÄUBLI LOUD

Swiss Federal Office of Public Health, Switzerland

Since evaluation systems are context dependent, they must take into account constituents' needs, wants, and expectations plus other variables such as pertinent societal values, customs, and mores; relevant laws and statutes; economic dynamics; political forces; media interests; pertinent substantive criteria; organizational mission, goals, and priorities; organizational governance, management, protocols, and operating routines; and the organization's history and current challenges. (Stufflebeam, 2002)

Introduction and Structure

Evaluation practice in Swiss public administration is still relatively new (Bussmann, 1996; Furubo et al., 2002; Jacob and Varone, 2002). Yet already, there are a number of indicators to suggest that it is gradually becoming 'institutionalized' at both cantonal and federal levels. There is an established, informal network of evaluation commissioners (*Netzwerk Evaluation in der Bundesverwaltung*) working in the federal administration that has met regularly for the last decade. A national evaluation society (Swiss Evaluation Society: SEVAL) of over 200 members exists that provides a regular newsletter, an annual conference, and, more recently, an approved set of Evaluation Standards pertaining to the quality and value of evaluations, as well as the expected professional conduct of the Swiss evaluation community (Swiss Evaluation Society, 2000). SEVAL also maintains a database of experienced evaluators who can offer a range of qualitative and quantitative methods, and from this winter, SEVAL is collaborating with the University of Fribourg to provide a series of short courses for the professional development of evaluators. For introductory courses, there are two on evaluation; one offered by the Continuing Education Department, Berne University and the other by IDHEAP (*Institute des hautes études en administration publique*) in Lausanne.¹ The fact that SEVAL has been concerned with codes of behaviour and evaluator competencies demonstrates its increased commitment to the development of good practice and professionalism.

However, one of the most significant indicators of the institutionalization of evaluation in the policy arena is the number of evaluation clauses that have been

introduced into legislation at both cantonal and federal levels over the past decade. A constitutional article (Article 170) was also recently introduced into the revised Swiss Federal Constitution (1999), which states that the 'effectiveness of federal measures must be the object of an evaluation'. Within the last 12 months, an inter-departmental working party (IDEKOWI) has been set up by the Federal Conference of General Secretaries to recommend strategies and procedures for the practical implementation of this article. Its report is due in 2004.

But 'institutionalization' does not necessarily equal 'integration'. Evaluation can be a threatening activity. Many groups and organizations struggle with how to build a good evaluation capability into their everyday activities and procedures. This is essentially an organizational culture issue. The IDEKOWI recognized from the onset that the effective and beneficial use of evaluations could not therefore be contemplated without the development of a so-called 'evaluation culture'. In line with the consensus nature of the Swiss political system, it is recognized that the attainment of a careful balance of 'top-down' and 'bottom-up' processes is necessary to achieve such a goal. It would not be constructive, for instance, to impose evaluations on the federal administration without gaining the interest and support for evaluation practice within the federal administration itself. Therefore 'bottom-up' experiences have also been integrated into this process: a representative from the federal administration's network of evaluation commissioners is also a member of the IDEKOWI.²

In this presentation, I shall draw on my experience within the Swiss Federal Office of Public Health (SFOPH) and as a member of the IDEKOWI to describe the uses and limitations of some of the means we have at our disposal (Guidelines and Standards) towards our common end (an evaluation culture).³ In particular I will argue that efforts towards developing an evaluation culture must include one vital ingredient: training.

Firstly, however, I shall set out a vision of an ideal 'evaluation culture', then explain the historical background to the development of both the SFOPH Evaluation Guidelines and SEVAL's Evaluation Standards to help clarify their uses and limitations as described in the next section. Finally I shall draw out some general lessons from the (quite limited) experience we have had so far.

What is an Evaluation Culture?

'Culture' has been defined (by *Collins English Dictionary*) as 'the total of the inherited ideas, beliefs, values and knowledge, which constitute the shared bases of social action' (Makins, 1995). The following paragraphs outline some of the essential characteristics I believe an evaluation culture should have (and I try to foster in my workplace). Most of these, and many more, can be found in the relevant literature and indeed, they mirror the basic principles of the SEVAL and other societies' evaluation standards.

First and foremost, the ideal 'evaluation culture' is characterized by a willingness to learn. It is made up of leaders and colleagues wanting to find solutions to problems, willing to try out new ideas and identify where improvements can be made. The context is 'safe' enough for constructive, not destructive criticism

to be voiced. The positive as well as the negative lessons are identified together as part of the learning process. Evaluations are seen and used as a means of 'improving, not proving' (Smith, 1998).

Self-evaluation within an organization is the norm; it is an ongoing process that is considered part and parcel of 'good management'. Self-criticism about the process and effects of its work is encouraged and praised. External evaluation specialists are commissioned for an impartial judgment of programmes, policies and measures that are strategically important and/or potentially controversial. Information provided through evaluation is highly valued and respected.

External evaluators also are self-critical and reflective about the procedures and methods used for their evaluations. They recognize and openly acknowledge the limitations of their work, and do not over-exaggerate or make inappropriate generalizations about the results.

The culture is action-oriented, ready to draw out and then use the evidence available to revise, reorient and/or transpose the lessons learned to other situations. Both the process (of being involved) and product (the evaluation report) of evaluation are a means of stimulating thought and are an aid to planning – especially on how to use the results to best effect. Equally, however, the consequences of proposed actions are carefully considered before any action is taken.

Evaluations are designed to provide feedback in a timely fashion; evaluators anticipate when and in what form relevant information can be provided to help the decision-making process. Evaluation 'purchasers' are transparent about their intended use and information needs, and, of course about their timetable for decision making.

Evaluators, 'purchasers' and commissioners alike recognize the limitations of evaluation. Evaluators are not expected to make decisions on behalf of others, although certainly they are expected to provide the information needed to help decision makers. They recognize, however, that evaluation results and recommendations are only one of the inputs that are used in decision making. Evaluation purchasers, on the other hand, take account of the findings and results of evaluations in their planning. There is transparency about both the process and product of evaluations.

The existence of an 'evaluation culture' therefore very much depends on the degree to which beliefs, values and knowledge are shared about the process and product of evaluation practice. Guidelines and standards can both contribute in their own way to spreading knowledge about professional conduct and overall quality measures needed to optimize the usefulness of evaluations. Ultimately, however, it also depends on the quality of the evaluations themselves and the degree to which standards are therefore understood and applied.

Quality Improvement Measures

This section aims to set the developments of our Guidelines and Standards in context. It will become clear that, despite the relatively rapid development of evaluation practice in Switzerland, there remain gaps in the quality of the work and/or an understanding of how best to use the results.

SFOPH's Evaluation Guidelines

The SFOPH has had an evaluation tradition since 1987. In its efforts to combat the HIV/AIDS epidemic, the Office identified the need to evaluate the effects and effectiveness of its prevention policy. Due to the sensitive and pioneering nature of its prevention work (e.g. needle exchange programme in prisons, medical prescription of heroin, etc.) it was agreed from the beginning that the credibility of the evaluations would be better assured by commissioning studies from external specialists. Specialists in social and preventive medicine were sent to the UK for special training in the evaluation of public health measures.

The benefits of evaluation were quickly recognized and, as a consequence, external evaluations were commissioned for other aspects of SFOPH work such as illegal and legal drug use.

But the Office lacked experience in the commissioning and management of evaluations and therefore an internal unit was set up within the SFOPH's prevention division (1992) to do just that. This led to a change of policy. A system of competitive tendering for evaluation mandates was introduced to foster the development of a critical mass of evaluation specialists; partnerships were forged with a range of different institutions – both private and university-based; and, for quality assurance, a set of procedures and standards for the Office was provided in the form of guidelines: 'Guidelines for Health Programme and Project Evaluation Planning' (SFOPH, 1997). They include a set of pull-out checklists to prompt thought at each stage of the cycle, from evaluation planning to dissemination and use of the results (and back into planning). They were principally aimed at programme and project managers, although they also include checklists for evaluators.

The Guidelines were originally intended to help managers prepare for their discussion with in-house evaluation experts when planning evaluation studies. They were to have been introduced through a series of training workshops; such training never came about due to lack of top management support and the streamlining of evaluation as part of the structural reorganization of the Office. The internal evaluation unit's staff resources were reduced, and it was re-located to another division with no experience or history of evaluation. With this change, the *intention* of the Guidelines also changed: managers were expected to use the Guidelines to plan and manage their evaluations without the support of the in-house evaluation unit. After three years of trial and error, this plan was abandoned.

With the introduction of Article 170 into the revised Federal Constitution, there was renewed interest at executive level in evaluation. An 'internal evaluation culture', it was said, was needed to assure the optimal use and utilization of evaluation. In 2001, an Evaluation Management and Resource Centre (*Centre de competences en evaluation: CCE*) was therefore set up at directorate level with five staff. It is charged with not only commissioning and managing the evaluation of priority policy, programme and measures, but particularly with fostering an 'evaluation culture'. The CCE has now developed the conceptual framework for an Office-wide internal evaluation system, and the procedures needed to assure its application (including a training programme for middle managers) are gradually being introduced.

To summarize, although evaluation has been around for many years in the SFOPH, as yet, there are limited signs of an established 'evaluation culture'. Whatever shared understanding there is of evaluation, particularly its role and contribution to policy, strategic and programme development, has been, and still is, very much limited to the few who have actually been involved in an evaluation. With no formal training programme yet in operation, in-house learning has essentially taken place 'on the job' with the support of the Guidelines and in-house expertise.

SEVAL's Evaluation Standards

The SFOPH Evaluation Guidelines were essentially conceived as a 'planning tool' for SFOPH managers (not evaluators) and were certainly not intended as a do-it-yourself methodological toolkit. They made no attempt to set down any rules or 'code of practice' in the conduct of evaluation. Fortunately, the SEVAL's Evaluation Standards were developed to fill this gap.

The procedures used to develop the SEVAL Evaluation Standards are described in the document itself (*SEVAL Evaluation Standards* [Swiss Evaluation Society, 2000]). I shall therefore merely summarize some of the key features of this process.

The need for establishing a set of standards was identified as a consequence of a national research programme (PNR 27 1987–92) on the use and application of evaluation. A meta-evaluation (Widmer, 1996) of some of the studies showed that the evaluations did not always meet the quality and expectations of the commissioners (and vice versa!). Further studies have confirmed these findings (Bergman et al., 1998); they found:

- little formal training in evaluation;
- self-study 'on the job' training;
- insufficient or inappropriate use of methods;
- a lack of co-operation either between evaluators themselves, or between evaluators and their clients.

In 1994 therefore, a working group was set up by the newly formed SEVAL to elaborate some quality standards. The author of the meta-evaluation study, Thomas Widmer of Zurich University, chaired the group. The membership totalled 15 and included commissioners (both cantonal and federal levels) and evaluators (consultants from private bureaux and university researchers).

The aim of the group was to produce a set of standards that, when applied, would help improve quality and professionalism within the evaluation community. After a review of what was on offer worldwide, the Evaluation Standards Working Group finally agreed to base its work on an adaptation of those produced by the American Joint Committee on Standards for Educational Evaluation.

A drafting committee was established and the work was concluded through an iterative process between this committee and the SEVAL Standards Working Group. A special SEVAL conference was then held on the subject (26 May 2000) to listen to, and take into account, members' comments before the final

document was prepared and submitted for approval to the SEVAL annual general meeting later that year. The final version is therefore based on another country's existing standards, but adapted to the Swiss federal context.

But even the best of plans do not always stand up to the test. How feasible and practicable are the Standards proving to be?

Use and Usefulness of the Quality Improvement Measures

Even though the Guidelines were not put to use under optimal circumstances, we have gained several years of experience about their use. As for the Standards, we are only just at the beginning of 'testing' their use and utility. The use and usefulness of both Guidelines and Standards are described below both from the perspective of the SFOPH's CCE and others.

Use and Usefulness within the SFOPH

Let me start with an overview of how we are now using the Office's Evaluation Guidelines and the SEVAL Standards within the SFOPH.

A copy of the Guidelines is provided for reference to middle managers, particular project managers, and new staff. The individual Checklists that accompany the SFOPH's Guidelines are systematically used by the CCE to commission and manage an evaluation. They are initially used as the basis of discussions with internal partners during the planning stage of an evaluation to identify the key evaluation questions as well as the scope and focus of the study. Similarly the checklists alert evaluators to our requirements for evaluation designs and reports.

The checklists help prompt thought and discussion about:

- why the evaluation is needed, for whom, and to what end;
- what needs to be included in the evaluation's Terms of Reference;
- the essential items that need to be addressed whilst managing the evaluation;
- what should be included in an evaluation design, technical report, the executive summary; and
- drawing out the key lessons and planning how best to use evaluation results.

In terms of quality assurance, the CCE has transformed the SEVAL Standards into a checklist and questionnaire for:

- judging the quality of the evaluation designs sent in response to evaluation bids; and
- the meta-evaluation of intermediate and final reports.

Meta-evaluations of commissioned evaluations are systematically carried out by CCE staff, who are responsible for quality assurance. Occasionally, for the more controversial studies, external experts are also called upon to conduct meta-evaluations. The questionnaire and checklist based on the SEVAL Standards are used systematically in both cases; line managers are prompted about the utility of the evaluation results by means of a modified form of the 'utility'

criteria from the SEVAL Standards. All CCE evaluation contracts now state that the quality of external evaluations, both process and products, are to be judged according to the criteria set out in the SEVAL Standards.

So far, both Guidelines and Standards are used by the CCE for the planning and utilization of individual external evaluation studies. For line managers, they have proven most useful when used as an aide-mémoire in planning discussions between managers and CCE staff. They focus on the precise points that need to be taken into account at each stage of the process and in a simple, easy-to-understand manner. Previous experience, just prior to setting up the CCE, showed that, when used independently, the Guidelines were not used to best effect.

Equally, the checklists are used by the CCE in negotiations with evaluators about its requirements. They provide a clear set of criteria against which to judge the quality of evaluation practice. The Standards can help identify the limitations of the study and/or where further knowledge/training is needed – for both evaluators and CCE staff! We are currently updating our complete set of Checklists to comply with the various points set out in the SEVAL Standards.

Use and Usefulness for Other Users

We know from our order database, that a wide range of institutions and individuals outside of the SFOPH has ordered the Guidelines. From the spontaneous responses we have had, it appears that they have been mainly used for reference in planning evaluations and for teaching purposes. However, we have made no attempt to systematically assess how they have been used and to what effect.

The wider use of the SEVAL Standards has recently been addressed in a newly formed SEVAL working group, which met regularly over the winter 2002–3 to develop recommendations for their promotion and use. Reported use to date included:

- in evaluation contract management, for meta-evaluations by evaluation commissioners within the federal administration;
- in teaching, for meta-evaluations by university students whose studies include modules on evaluation of public administration and/or policy measures;
- by evaluation consultants in supporting self-evaluations – particularly with teachers in the formal education sector.

However the group were not advised about how evaluators are using or could use the Standards for judging their own work. This is an important use about which we unfortunately know very little. At present, the Standards appear to be very little known outside of the SEVAL members as their promotion and use are essentially limited to a handful of individuals.

The group concluded that the SEVAL Standards were a potentially useful tool for promoting quality in evaluations. They could be applied to a range of different domains, and would be particularly useful to evaluation purchasers/commissioners. However, they were thought to be of very limited use for guiding novices in the practice of ‘self-evaluation’.

In conclusion the group recommended that the Standards should be systematically adopted and used by evaluation commissioners as a contractual requirement for quality assurance. The IDEKOWI and the *Netzwerk Evaluation in der Bundesverwaltung* were identified as the two most important groups to target for actively promoting the SEVAL Standards.

As they stand, however, it was agreed that the Standards are difficult for the novice to evaluation to interpret, and even more difficult to apply; a sound, understanding of evaluation principles and methodology is a prerequisite. The group therefore recommended that a set of checklists should be developed to accompany the Standards, together with case studies to illustrate the different points. Equally, and in line with the SFOPH's own conclusions, it was recommended that training civil servants (and evaluators) in evaluation and the use of SEVAL Standards was also vital for assuring quality evaluation.

What We Have Learned So Far

Guidelines and standards can contribute to assuring quality and credible evaluations. They provide a sound basis for quality assurance. However, not in isolation; they need to be used as part of an overall strategy for promoting and developing an evaluation culture. But from our limited experience to date we have learned that an 'evaluation culture' very much depends on educating evaluators and commissioners alike about the characteristics of quality evaluation and quality utilization.

In themselves, guidelines and standards are of limited use; the concepts behind them have to be explained through training. Yet at the same time, their systematic application through meta-evaluations can also help identify the knowledge gaps and consequently, training needs.

First and foremost, however, an 'evaluation culture' very much depends on recognizing the benefits and limitations of evaluation research and how it can contribute to improving policy and performance. Case studies of good and less-good practice need to be integrated into training towards this end.

Notes

An earlier version of this contribution was presented to the 5th European Conference on the Evaluation of Structural Funds: 'Challenges for Evaluation in an Enlarged Europe', 26–7 June 2003 in Budapest, Hungary.

1. See www.swissuni.ch for information on these courses, and the professional development courses at Fribourg University.
2. The Swiss Federal Office of Public Health, being one of the two Federal Offices with the longest tradition in evaluation, was elected by the *Netzwerk Evaluation in der Bundesverwaltung* to represent its membership in the IDEKOWI.
3. Definitions:

guideline: a procedural suggestion intended to help evaluators and their audiences to meet the requirements of the evaluation standards; strategy to avoid mistakes in applying the standards;

Läubli Loud: Setting Standards and Providing Guidelines

standard: a principle commonly agreed to by experts in the conduct and use of evaluation for the measure of the value or quality of an evaluation.

References

- Bergman, M., S. Cattacin and M. Läubli Loud (1998) *Evaluators Evaluating Evaluators: Peer-assessment and Training Opportunities in Switzerland*. Berne: Swiss Federal Office of Public Health. Available on: www.health-evaluation.admin.ch (site visited: 23 March 2004).
- Bussmann, W. (1996) 'Democracy and Evaluation's Contribution to Negotiation, Empowerment and Information: Some Findings from Swiss Democratic Experience', *Evaluation* 2(3): 307–19.
- Furubo, J.-E., R. Rist and R. Sandahl (2002) *International Atlas of Evaluation*. New Brunswick, NJ/London: Transaction.
- Jacob, S. and F. Varone (2002) *L'institutionnalisation de l'évaluation des politiques publiques: expérience étrangères et scénarios pour la Belgique [Institutionalization of the Evaluation of Public Measures: Other Countries' Experiences and Options for Belgium]*. Third intermediate report of the research project AM/10/016: *L'évaluation des politiques publiques en Belgique [Evaluation of Public Measures in Belgium]*. Belgium: Université catholique de Louvain.
- Makins, M., ed. (1995) *Collins English Dictionary*. London: HarperCollins Publications.
- Smith, N. L. (1998) 'Designing Investigative Evaluations: Problem Solution Versus Client Responsiveness', *Evaluation* 4(2): 117–29.
- Stufflebeam, D. L. (2002) *Institutionalizing Evaluation Checklist*. Kalamazoo, MI: Western Michigan University. Available at: www.wmich.edu/evalctr/checklists (site visited: 23 March 2004).
- Swiss Evaluation Society (2000) *SEVAL Evaluation Standards*. Available at: www.SEVAL.ch (site visited: 23 March 2004).
- Swiss Federal Constitution (1999) *Article no. 101.4*. Berne: BBL. (Revised 18 April 1999.)
- Swiss Federal Office of Public Health (1997) *Guidelines for Health Programme and Project Evaluation Planning*. Berne: Swiss Federal Office of Public Health. Available at: www.bag.admin.ch/cce/publikationen/leitfaden/e/leitfaden.pdf (site visited: 23 March 2004).
- Widmer, T. (1996) *Meta-Evaluation: Kriterien zur Bewertung von Evaluationen*. Bern: Paul Haupt.