

Embedding Evaluation in the Swiss Federal Administration

Purpose, Institutional Design and Utilization

THOMAS WIDMER AND PETER NEUENSCHWANDER
University of Zurich, Switzerland

Scholars have long addressed the question of how to improve the usefulness of evaluation in the public sector. This article describes the role of evaluation within the Swiss Federation in order to investigate the conditions of evaluation use. The article is based on an examination of several offices in the federal administration. Evaluation within the Swiss federal administration shows great variety among the different units with regard to both the general understanding and the implementation of evaluation. These provide good examples for studying the relationship of purpose, utilization and institutional design in evaluation. Furthermore, decisions are often taken in the absence of basic considerations regarding the purpose or utility of the evaluation, and this substantially affects the subsequent institutional design of the evaluation. The authors recommend a clearer functional differentiation between different forms of evaluations as means for improving the evaluation measures currently used in and by the Swiss federal administration.

KEYWORDS: evaluation purpose; evaluation utilization; institutional design; Swiss federal administration; Switzerland

Introduction¹

Great variety can be observed in the way in which evaluations are institutionally embedded in the Swiss Federal administration, and this variety can be ascribed to the highly fragmented structure of the administration. Some parts of the administration, however, have shown themselves to be far more committed to utilizing evaluations than others, and this article examines which specific institutional settings have shown themselves to be more favourably inclined toward evaluation. The variety within the Swiss Federation is well suited for studying the relevance of institutional settings in examining the utilization and purpose of evaluation. One aim of this article is to identify the predisposing factors, another is to better link intent with use. Unless we understand what function evaluation is to serve, we cannot properly assess the way in which evaluation is used in the administration.

The authors argue that the link between evaluation purpose and evaluation utilization plays a crucial role in defining the ideal institutional setting for evaluation. The article is structured as follows. In the next section, the culture of evaluation at the Swiss federal level is described, followed by a more theoretical discussion of the concepts of evaluation purpose and evaluation utilization. We then present two fundamentally different forms of institutional settings for evaluation within public administration and their relevance for purpose and utilization. In the following section, the empirical investigation of practices in the Swiss federal administration is discussed. Based on these findings, the article concludes with recommendations for a successful integration of evaluation purpose and utilization in institutional design.

The Evaluation Culture within the Swiss Political System

The evaluation culture in Switzerland does not have a long tradition in comparison with other countries (Furubo et al., 2002, esp. Spinatsch, 2002; Horber-Papazian and Thévoz, 1990: 133–43). According to Leeuw (2000: 60), Sweden, Germany and the UK are considered the European front-runners in establishing cultures of evaluation. Other countries, particularly in Scandinavia and Catholic Europe, have recently also made progress in developing a more thorough evaluation culture. Although Switzerland does not appear on such lists, a growing evaluation culture has established itself in recent years. Swiss evaluation was initially influenced more by regional economic concerns and political science and less by educational theory and psychology, the latter having been more influential in the US (Bussmann, 1995: 88).

One of the most important starting points for the development of evaluation activities was the Working Group on Legislative Evaluation (AGEVAL: *Arbeitsgruppe Gesetzesevaluation*) which was set up in 1987 (Bussmann, 1995, 1996). AGEVAL was composed of federal and cantonal officials as well as academic scholars, and it promoted evaluation through a series of commissioned evaluation studies. A second important milestone – rooted in the work of AGEVAL – was the National Research Programme 27 entitled ‘The Effectiveness of Public Measures’. This research programme was financed by the Swiss National Science Foundation, the national body for the promotion of scientific research in Switzerland. One of the main goals of this programme was to professionalize the then weakly developed evaluation community in Switzerland.

In 1996 – roughly a decade after AGEVAL began, and in essence a direct outcome of the National Research Programme 27 – the Swiss Evaluation Society (SEVAL) was founded. This association currently has about 300 members and brings together all those concerned and involved with evaluation, whether in politics, administration, the private sector or in academe, through its scholarly journal (*LeGes – Gesetzgebung & Evaluation*) and regularly published newsletter, as well as its yearly conferences. It also hosts special events focusing on selected themes, has established various working groups to address specific issues and evaluation fields, and has a recently updated website. With its adoption of evaluation standards in 2001, SEVAL has also articulated a new professionalism

for those who engage in evaluation (Widmer et al., 2000). In essence, a strong evaluation community has emerged in Switzerland over the last few years, so developments on the supply side seem promising. But what is the situation on the demand side?

Since 1999, the Swiss Federal Constitution has included a general evaluation clause in Article 170 requiring the Federal Parliament to ensure that the effectiveness of federal measures be investigated. Switzerland is – as far as the authors know – the only country worldwide with a constitutional article on evaluation. The demand for evaluation has increased significantly over the last decade as a concomitant to various structural reforms of the state (at all levels), and the steering of public bodies is legitimized more and more not only by input (legality and democracy) but by output (performance) as well. These developments have led to an increasing demand for evaluative services in Switzerland.

Despite these developments, Switzerland still ‘lacks a strong evaluation community’ (Widmer, 2000: 69; Spinatsch, 2002) due to weaknesses on the demand side that derive from several characteristics of the political system. One of these is the strong federalism that exists in the Swiss Confederation (the official name of this nation state) that functions at communal, cantonal and federal levels (Horber-Papazian and Thévoz, 1990; Bussmann, 1995). The 26 cantons (sub-national territorial units) have considerable autonomy, as each have their own constitutions, parliaments, executives and courts and each canton contains communes (municipalities) of which there are about 2900. An important organizing principle, strongly related to federalism, is subsidiarity, which means political tasks are entrusted to the lowest possible political level. Thus, the tasks at the federal level include foreign and security policy, and customs and monetary affairs, but public health or education remain the domains of the cantons. The communes have their own competencies in areas such as social affairs, the energy supply, road building or local planning, but because each canton specifies the degree of autonomy its communes can enjoy, there is considerable variation at this level. Thus, a very large part of state activities remain at the lower political levels of commune and canton, and the cantons are also responsible for implementing most federal legislation. Therefore, in addition to the Federation, cantonal administrations and executives commission many evaluations. The highly decentralized structure results in a relatively weak federal administration, and its lack of resources restricts the capacity of the federal level to conduct or commission evaluations (Fornerod, 2001: 29). For the same reason, cantonal and communal bodies often lack the critical mass to conduct evaluation in a professional way.

Various scholars (Spinatsch, 2002: 375–6; Derlien, 1997; Bussmann, 1995: 96) have suggested that another reason for the weakness of the evaluation community is that Switzerland practises a form of semi-direct democracy: a wide range of direct democratic instruments exist at all political levels, the most important of which at the federal level are the popular initiative and the referendum (Linder, 1994: 84). Although the success rate of popular initiatives is small, the indirect effects can be substantial, inasmuch as initiatives exert pressure at the federal level to move forward and undertake major or minor legal

and administrative alterations in response. Referenda, by contrast, can block or delay laws passed by Parliament or actions taken by the executive, and are thus often considered to be similar to a veto. Bussmann states: 'When evaluations produce results that conflict with decisions sanctioned by the electorate, it is quite likely that the former will be shelved' (1995: 95). Fornerod (2001: 29) claims direct democracy has an influence on evaluations, and can act either as a brake or as a catalyst.

A third characteristic of semi-direct democracy is the strongly consensus-oriented political mode. To avoid failures in subsequent plebiscites, administration, government and parliament strongly emphasize the need to ensure broad support prior to suggesting any new policy. Therefore policy making in Switzerland is oriented toward gathering support more than toward gathering evidence. Only under specific circumstances does empirical evidence play a crucial role in policy formulation in Switzerland.

Overall, the picture is contradictory. A positive development on the supply side, and in part on the demand side, is hampered by some of the specific aspects inherent to the Swiss political system. As will be discussed below, this favours diversity in evaluation at the federal level, thereby providing rich empirical evidence for various evaluation practices. But first we turn to the conceptual framework for studying these practices. In the next section, the various purposes and uses of evaluation are discussed. We then relate purpose and utilization to the institutional design in which evaluations are carried out.

The Purpose and Utilization of Evaluations

The literature describes different purposes for evaluations (Scriven, 1991; Weiss, 1989, 1995: 145–6, 1998: 20–45). Vedung (1997: 101–14) suggests there may be four general purposes evaluations can serve: accountability, improvement, basic knowledge, and strategy. In *accountability*-oriented evaluations 'information and evaluative judgments are produced to allow decisions on program continuation, expansion, reduction, and termination' (1997: 102). If *improvement* is a goal, 'evaluation aspires to guide program amelioration and refinement' (1997: 108). *Basic knowledge* evaluation, on the other hand, is seen 'as fundamental research that seeks to increase the general understanding of reality' (1997: 110). Finally, if *strategic* purposes dominate, then evaluations are used 'to hide shortcomings and failures from their principals, to display attractive images of programs, and in general to provide appearances more flattering than reality' (1997: 111). These different evaluation purposes have an influence on the quality of evaluation, on the utility as well as on the evaluation-specific institutional design. On the other hand, the organizational design also exerts an influence on the utilization of evaluation findings and recommendations.

Evaluation utilization has long been one of the most prominent issues in the international evaluation literature, and the implementation of evaluation results and recommendations was often measured by the percentage of recommendations that were implemented (Owen and Rogers, 1999: 108). This kind of evaluation utilization is often designated as instrumental use (Mayne, 1994: 19).

Yet the research also shows that evaluation studies frequently are ignored or politicized (Patton, 1997: 7), and not only in the US or the UK, but also in Switzerland. Zimmermann et al. (1989), for example, discovered that out of 77 evaluation studies executed at the national level, 57 percent did not have any influence on politics, while 22 percent led to a broader public discussion and only 21 percent of the examined evaluations were implemented (as self-reported by evaluation authors). Matters have supposedly improved since then, however, and at least according to Fornerod (2001: 77), from 20–50 percent of federal evaluation studies are used extensively or very extensively (whatever this means) in the political decision-making process. Only 5–10 percent of all evaluation studies had no effect. Indeed, evaluations often have uses the effects of which are not directly measurable (see Mayne, 1994: 20–1). A similar viewpoint is put forward by Carol Weiss, one of the best-known researchers in this field, who points to the learning effects from evaluations that may lead to:

- a better knowledge of the effects of measures;
- changing priorities;
- the invention of new ideas; or
- a decrease of uncertainty without resulting in immediate decisions or new measures (Weiss, 1980).

Weiss coined the term ‘enlightenment function’ to describe such learning effects (Weiss, 1977). In various contributions, Weiss has shown that thanks to evaluations, stakeholders can better recognize and understand policy and implementation problems (Weiss, 1980, 1981, 1988).

A framework is employed here – also proposed by Vedung – that goes beyond the differentiation between instrumental and conceptual or enlightenment uses. This framework sums up various concepts put forward in the broad literature on evaluation utilization and allows for a comprehensive view of various forms of utilization. Vedung describes five uses for evaluation: instrumental, conceptual, interactive, legitimizing and tactical (1997: 265–89). If ‘evaluation findings are utilized as means in goal-directed problem solving processes’, this is an *instrumental* use (1997: 269). *Conceptual* use occurs when cognitive and normative insights are gained through evaluations, ‘but these insights have not been directly instrumental and transformed into action’ (1997: 272). The procedural use of evaluation findings by various participants, including journalists, political parties or other interests, is designated *interactive* use, though the ‘process from evaluation to future decision in the political system is not linear and not unidirectional but unorganized, messy and interactive’ (1997: 274). If evaluation ‘is asked [. . .] not to meet a genuine interest to know more about the world but to use as ex post rationalization, to legitimize decisions and resolutions based on other considerations’, the role of evaluation may be a *legitimizing* one (1997: 275). Finally, it may be a *tactical* use when ‘evaluation is requested to gain time or avoid responsibility’ (1997: 276).

Evaluation-specific Institutional Design in the Context of Purpose and Utilization

Evaluation-specific institutional arrangements can take many different forms. Two ideal-type organizational models have been developed – centralized and decentralized (see Figures 1 and 2 below; cf. Williams et al., 2002: 43–6) – to aid the systematic discussion of the evaluation arrangements observed. In the centralized model, evaluation responsibility is assigned to a specialized evaluation unit at the division level. This evaluation unit suggests potential evaluanda and establishes accompanying advisory committees that include members of several internal sections or occasionally other offices or external experts. The specialized unit commissions an external evaluator and is responsible for the dissemination of evaluation results and recommendations. The planning of evaluation programmes often needs approval by a directorate (senior policy/programme managers). As the main input for carrying out evaluation comes from units at the directorate or divisional level, the centralized model can also be described as a top-down model.

The second type of evaluation arrangement is decentralized. In this case, the units in charge of implementing evaluations are the various sections within an office. The section in charge of the respective policy suggests potential evaluanda

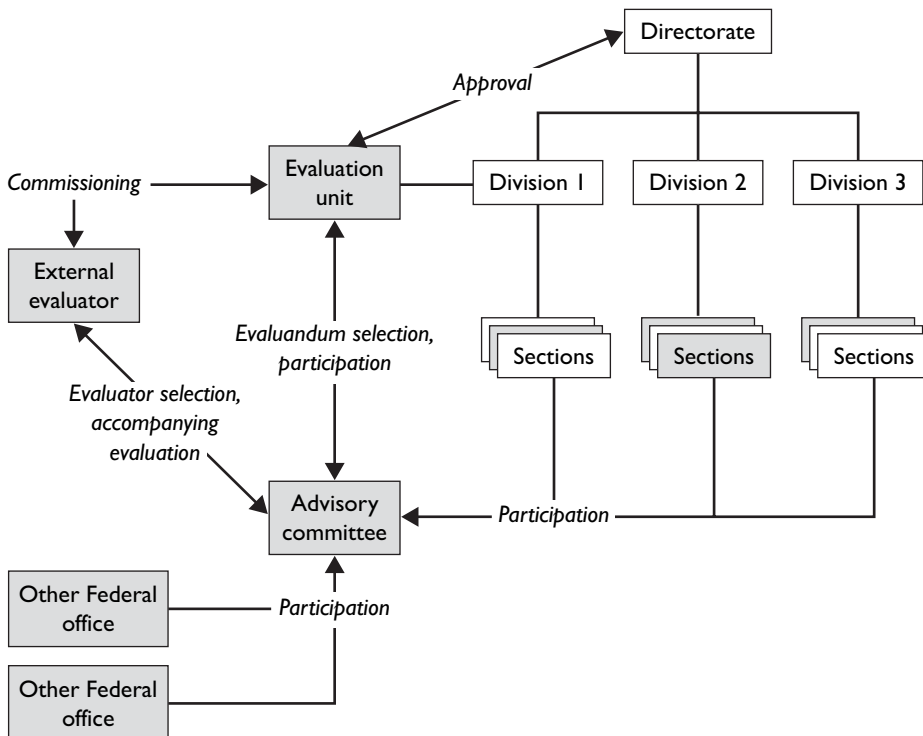


Figure 1. Centralized Evaluation Setting

and a specialized evaluation unit may provide support for their evaluation efforts. The specialized evaluation unit is normally part of the staff of the administrative body without independent decision-making power. As the main inputs for evaluation execution come from low-level administrative units, this model can be described as bottom-up. The directorate is not directly involved in the evaluation process, so its approval for a planned evaluation is usually unnecessary. Furthermore, the evaluation unit has less influence on the evaluation than in the centralized model.

It is obvious that the decentralized type fits an evaluation whose purpose is improvement far better than does the centralized type, which is in turn more appropriate for an accountability-oriented evaluation. Furthermore, the decentralized setting is also more appropriate if one has an instrumental or interactive utilization in mind, whereas the centralized configuration is better at providing conceptual, legitimizing, and tactical uses (Sonnichsen, 1999: 64; Patton, 1997; Weiss, 1998: 39–40; Widmer, 2002).

In Table 1, the different forms of purpose and utilization are related to their most appropriate institutional setting. This framework assumes that:

- we do not expect to observe the basic knowledge purpose (at least as a primary concern) in an applied setting (marked with an N);
- we rarely expect to observe strategic purpose and tactical use, as these forms are rarely mentioned in official documents or by officials – they are assessed as not legitimated (marked with an N);
- conceptual use is very hard to observe in an empirical way (marked with an N);

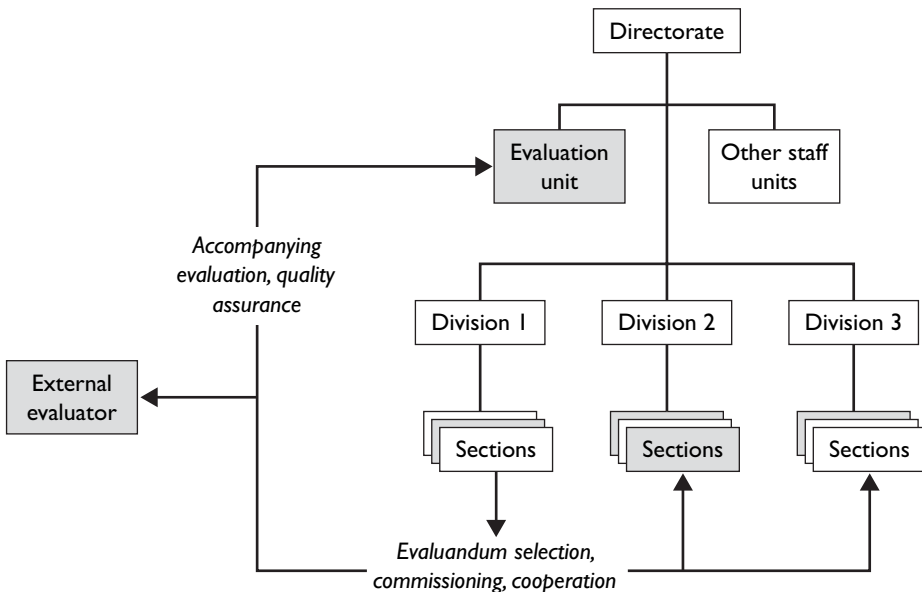


Figure 2. Decentralized Evaluation Setting

Table 1. The Most Appropriate Institutional Setting in Relation to Purpose and Utilization

Purpose	Utilization				
	Instrumental	Conceptual	Interactive	Legitimizing	Tactical
Accountability	M	N	M	C	N
Improvement	D	N	D	M	N
Basic knowledge	N	N	N	N	N
Strategy	N	N	N	N	N

Notes:

C: centralized; D: decentralized; M: misfit of purpose and utilization; N: not expected to be observed in an empirical investigation

- some combinations of purpose and utilization do not fit together, as in an evaluation planned with an accountability purpose with an interactive use (marked with an M);
- centralized or decentralized settings are more appropriate for specific purposes and uses (marked with a C or a D).

This framework allows us to systematically describe the various configurations observed in evaluation within the Swiss Federation. In the next section these configurations are described for eight federal offices. Each description is concluded by an assessment of the three variables purpose, utilization and institutional design.

Embedding Evaluation Activities within the Federal Administration

Swiss Agency for Development and Cooperation

Evaluation within the Swiss Agency for Development and Cooperation (SDC) has a tradition that stretches back 30 years. Today evaluation functions are embedded in Programme Cycle Management (PCM; cf. Eggers, 2002), a system that is both result- and process-oriented. PCM allows stakeholders to improve the quality of their activities through joint planning, implementation, monitoring, and (self-) evaluation in the context of joint activities. The SDC differentiates between two forms of evaluation: self-evaluation and external evaluations. The SDC also distinguishes between evaluation objects. Project evaluations constitute the great majority of all evaluations within the SDC. These short-term evaluations take two to four weeks and usually cost less than SFr. 50,000. Evaluations of projects or programmes that belong to a specific sector in one or more countries are called sector evaluations. The strategy of Swiss development and co-operation in a specific country, pursued in co-operation with the Directorate of Political Affairs and the State Secretariat for Economic Affairs, is subjected to country programme evaluations. Overarching topics, such as poverty, healthcare or job training are examined in cross-sectional analyses that are both demanding and costly. Five thematic analyses were commissioned in the early 1990s.

The main responsibility for the evaluation function lies in the sections or in the co-ordination offices outside of Switzerland. The questions to be addressed in the evaluations are decided in the sections in conjunction with the units for evaluation and control situated in the five SDC departments. This evaluation system was sharply criticized by the OECD's Development Assistance Committee. It pointed out that 'the fact that the evaluation units are part of the SDC is not conducive to their independence.' Another point of criticism was that 'evaluations tend to focus on the implementation of the different parts of a project rather than on their results and impact' (OECD's DAC, 2000). The decentralized responsibility also results in evaluation reports that differ greatly in quality between the different sections and divisions. SDC officials are aware of this criticism but are convinced that this learning-oriented evaluation arrangement improves the implementation of evaluation results.

Aside from the five evaluation and control units at the divisional level, there is an additional evaluation unit in the Directorate. This unit is responsible for the overall planning, organization and execution of evaluations carried out by external evaluators, and also systematizes results and disseminates lessons learned.

The SDC management is not informed about the specific results of a project evaluation. However, it does receive an annual overview of the evaluation activities as well as continuous abstracts of completed evaluation studies. When the results of extensive evaluations exist, they are promoted through special presentations to which management members are also invited. Nevertheless, such broad dissemination of evaluation results is still exceptional in the SDC. In addition, evaluation reports are usually not published, but can be ordered on request. SDC officials consider many evaluation reports to be quite technical and difficult to understand and therefore do not want to put effort into publishing them.

To conclude, both of the ideal-typical evaluation arrangements exist within the SDC. The arrangement for widespread project (internal) evaluation can be described as decentralized because potential evaluations are decided in the sections (together with the decentralized evaluation and control units). For independent, external evaluations, the evaluation-specific institutional design is centralized because a high-level unit (the evaluation and control unit in the Directorate) is responsible for planning, organizing and executing this kind of evaluation. In this context, line staff play a supporting role in the preparation and execution of evaluations.

In accountability-oriented external evaluations (which follow the centralized model) the most important aspects of evaluation utilization are instrumental and legitimizing. Because of the public debate about external evaluation results, another dimension of such use is its interactivity. By contrast, the purpose of internal evaluation studies (which follow the decentralized model) is improvement and they are used in an instrumental way.

Federal Office of Public Health

Evaluation began to be a subject for the Swiss Federal Office of Public Health (FOPH) with the emergence of AIDS in the late 1980s. After establishing it in the field of AIDS, the office began to use evaluation in other domains as well.

Nowadays, evaluation within the office is mainly initiated in the 'Substance Abuse and AIDS' domain. The different sections and units ('Policy and Research', 'Drugs', 'Alcohol and Tobacco', 'AIDS', 'Campaigns and Marketing Section', 'Migration and Health', 'Health Promotion and Disease Prevention') are responsible for ensuring that their policies are evaluated. In 2000, FOPH's evaluation unit was affiliated to the specialist unit 'Public Health', but since September 2001, the so-called Evaluation Specialist Centre is now affiliated with the Chief of Staff of the Directorate. The evaluation unit is responsible for evaluation contracts, methodology and scientific quality assurance, though it usually does not propose subjects or themes for new evaluation studies.

The utilization of evaluation results and recommendations within the FOPH is highly dependent on the needs of the respective sections and on the context of the evaluandum. It is largely the decision of the head of the section what to do (or not to do) with the evaluation results. This practice carries the risk that evaluation results remain at the operational level and no one outside the section takes notice of the findings. Hence, the evaluation unit considers it important that results become part of strategic and political considerations. Therefore the unit helps involved sections explain the value and importance of evaluation findings, and tries to actively promote and facilitate productive use of evaluation findings at the Directorate level.

Complex and demanding programmes need to be implemented within the specialist unit 'Substance Abuse and AIDS' due to the policy issues raised by these public health concerns. Thus evaluation is an important control device or external assessment tool of how programmes are implemented (process evaluation) and of the policy outcomes of projects or programmes.

The evaluation-specific institutional design within the FOPH can be described as a decentralized evaluation arrangement. The decentralized sections decide if an evaluation is to be commissioned. The evaluation unit supports the sections and is also responsible for ensuring evaluation quality. The most important purpose evaluation serves is to improve projects and programmes. Thus, evaluation results and recommendations serve to incrementally adjust existing projects or programmes, and can therefore be described as instrumental utilization. But the results and recommendations also serve as legitimation: Switzerland pursues a progressive drug and AIDS policy, and results of accountability-oriented evaluations are used to justify the measures that have been taken in a controversial policy realm.

Federal Social Insurance Office

According to Federal Social Insurance Office (FSIO) administrators, research and evaluation in social security do not have a long-standing tradition in Switzerland. However, FSIO officials believe evaluations can offer a more systematic approach, give a more profound insight, and deliver better information about programme effects and effectiveness.

Within the FSIO, the specialized unit 'Economy, Foundations, Research' is responsible for evaluations. Of the projects this unit co-ordinates, 70 percent are evaluation studies. The specialized unit does not carry out evaluation studies but

Evaluation 10(4)

contracts out evaluations to external partners. Evaluation studies usually cost between SFr. 40,000 and 120,000, and they ordinarily last from 6 to 15 months. The current focus of the unit is to analyse the impact of a new health insurance law. Evaluation studies of the disability insurance and the national old-age pension schemes are planned for the near future.

The specialized unit prepares an outline in advance that describes the content, methodology, time needed, and budget proposed for each of the evaluation projects. After that, an advisory committee is formed, headed by a member of the specialized unit or section staff. An advisory committee usually consists of 3–8 members of the specialized unit or other FSIO sections, and may include officials from other offices as well (e.g. Federal Office of Public Health, Federal Statistical Office).

All FSIO directors receive the preface, synopsis and abstract of each evaluation report for comment, and may receive the complete evaluation report on request. The most important knowledge transfer occurs via the members of the advisory committee. Evaluation reports are sent primarily to cantonal administrators but also to the media, to research institutions, and to (non-)parliamentary commissions that work on the topic; 400–600 copies of the evaluation reports are usually published.

The specialized unit ‘Economy, Foundations, Research’ initiates FSIO evaluations and its scope and responsibilities are extensive. Therefore, this arrangement can be described as centralized. On the one hand, FSIO evaluations assess diverse programmes (for accountability purposes) and are used in a legitimizing way. On the other hand, evaluations within the FSIO are commissioned to improve public policy and are used in an instrumental manner.

Federal Office of Justice

The Federal Office of Justice (FOJ) executes evaluation mainly in two domains: help for crime victims; and correctional, rehabilitation and detention measures. Three evaluation reports were commissioned under the evaluation clause that was included in the Federal Law on Aiding Crime Victims. An advisory committee was formed that included members from inside and outside the FOJ. FOJ evaluations generally cost SFr. 60–70,000 each and take 4–5 months to complete. The FOJ also began a systematic investigation of the Federal Law on Aiding Crime Victims. As an expert commission was created in 2000, charged with the task of drafting a revised law, evaluation of assistance for the victims of crimes was a central basis for preparing the revision.

The other domain in the FOJ where evaluations are conducted is for correctional, rehabilitation and detention measures. In this area, the Federation supports new correctional measures carried out on the cantonal level, and evaluation is an integral and mandatory part of each of these pilot projects. The cantonal project leaders responsible for the pilot projects are the main clients of the evaluations. The FOJ’s role to ensure quality is maintained in the intermediate and final reports.

The procedure for pilot projects is quite a complicated one (see Figure 3). First, an application for a pilot project (including the evaluation concept) is

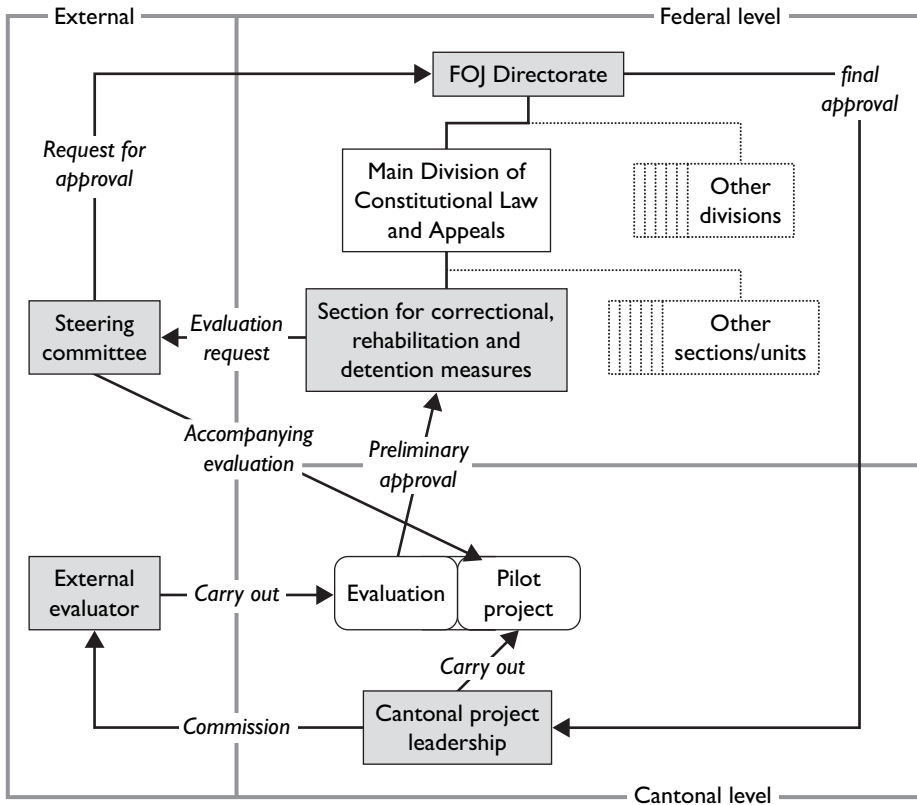


Figure 3. The Two-level Evaluation Setting of the Federal Office of Justice (Correctional, Rehabilitation and Detention Measures)

checked for its completeness and quality by the FOJ section responsible for it. After this first examination, proposals are sent to an external steering committee that consists of experts in correctional, rehabilitation and detention measures, and they in turn consider the pilot project’s significance, the budget and the practicality of the evaluation concept. After this preliminary verification, the steering committee requests approval from the FOJ to conduct the pilot project. A designated member of the steering committee accompanies the pilot project while it is being conducted, and reports her observations on behalf of the whole committee. When a final evaluation report is approved by the steering committee, the responsible FOJ section drafts an abstract, and all abstracts are published once a year in an FOJ information bulletin. The public can order evaluation reports as well, but demand for them is low. It is unknown within the FOJ to what extent, or even if, cantonal pilot projects leaders use these evaluation results.

Figure 3 shows that evaluation-specific institutional design in the FOJ ‘correctional, rehabilitation and detention measures’ domain does not fit either of the two ideal-typical arrangements. In contrast, the evaluation arrangement in the

Evaluation 10(4)

'help for crime victims' domain can be regarded as centralized. The central purpose of the three executed evaluations was accountability, and they were used in an instrumental manner.

Federal Finance Administration

Evaluation in the Federal Finance Administration (FFA) takes place within the scope of the Law on Subsidies that was enacted in 1990, a law aimed at standardizing federal subsidy practice. According to this law, government has to check periodically (at least every six years) whether the provisions on financial aid and compensation correspond to the Law on Subsidies, and a report has to be completed. A first part of the 'subsidy report' was published in mid-1997, the second part two years later.

The 'subsidy report' was elaborated in-house without extra resources and without extra support. Conceptual work was undertaken by the 'Expenditure Policy' division, and a project team was assembled to handle the task. Its members included representatives from each of the five financial services in the FFA, together with one delegate from the 'Finance and Accounting Division', one from the 'Financial Statistics and Equalization' section, and one from the Federal Audit Office. The project team was headed by the director of the 'Expenditure Policy' division for the first part of the report and by the chief of the 'Financial Service I' section for the second part of the report.

The first aim of the project was to obtain an overview of the subsidies that exist in the federal administration – this was important due to the great complexity in the system – and to this end a database was established that eventually contained information on a total of 660 different subsidies. About half of these subsidies were examined and assessed. As a consequence of these examinations, many individual reform measures were proposed. Since then, the FFA follows up subsidies annually.

Evaluation is carried out because of an evaluation clause in the relevant law. The purpose of evaluations within the FFA is therefore accountability. The type of utilization is instrumental and the evaluation setting is clearly a centralized one.

State Secretariat for Economic Affairs

Evaluation in the State Secretariat for Economic Affairs (Seco) takes place in the two sectors 'Labour Market Policy' and 'Growth Policies and Structural Reform'. The following considerations are confined to the 'Labour Market Policy' sector, for which the Swiss Federal Council requested evaluations. Due to the second revision of the 'Unemployment Insurance Law', the Swiss government changed the paradigm it used in labour market policy from passive to active policy measures. New instruments were created, including Regional Job Placement Centres as well as programmes for job creation and for retraining workers. One consequence was that a large administrative apparatus was created in a short time; another was that very high costs were incurred.

The government request for evaluation was split into five partial projects commissioned from academic and private institutions. In the evaluation arrangement, the 'Labour Market Policy' sector served as a link between the external

evaluation teams and the units in charge within the office. For each of the five evaluations, an advisory committee was established, its members including representatives from the 'Labour Market Policy' sector, employer organizations and the unions. Cantons, the main actors in implementing the new labour market measures, were also represented. Seco members who had special expertise in statistical data also provided support. The main reason for establishing an advisory committee was to provide the evaluation teams with relevant information as well as to guarantee that the evaluation studies would remain practically oriented. Methodological support was not relevant for the group given Seco's own expertise.

The results and recommendations of the evaluations were used to improve the different measures and instruments (e.g. on behalf of the Regional Job Placement Centres). The evaluation had a legitimizing character in part, because questions about the effectiveness of the costly new labour market instruments had been raised, and the evaluation results were able to respond to these questions. Another consequence was that an evaluation culture was established in the labour policy domain. Furthermore, due to the evaluations carried out, an effectiveness-based agreement with the cantons could be established. The five evaluation reports were completed in about two years and were subsequently published in a widely read journal, *Die Volkswirtschaft*. Still, the hopes placed in the tool of evaluation were at times too high, and for some the results proved disappointing.

The description of the Seco evaluation arrangement above suggests a highly centralized institutional setting. The most important purpose of evaluations is accountability as new labour market policy measures were introduced in Switzerland. Results were used in both legitimizing and improvement ways. Results were also utilized in an instrumental manner.

Federal Office for Agriculture

Internal and external evaluations are carried out in and by the Federal Office of Agriculture (FOAG). The evaluation arrangement in the FOAG is unique because six Research Stations are assigned to the Office, and they are available as evaluation partners. Although all interviewed officials admitted that this arrangement could cause problems, inasmuch as they were not structurally independent of the FOAG, this problem was seen as negligible in practice.

The role of evaluation within the FOAG grew historically. Although there is a central unit called 'Research Staff', evaluations are carried out in a decentralized manner. Each of the three divisions ('Production and International Affairs', 'Direct Payments and Structural Improvements' and 'Research and Extension') is responsible for organizing and executing evaluations on their own. Evaluation ideas are usually born in the different departments and sections (bottom-up). In some cases, 'Senior Management Staff' suggest evaluations, mainly for questions that are of higher political importance. These suggestions are bundled in the divisions and arrive, via the 'Division of Research and Extension', at the management level where decisions are taken.

One of the most important evaluation programmes within the FOAG is the

Evaluation 10(4)

evaluation of agricultural policy measures. Three FOAG units co-ordinate this evaluation in a collaborative manner. The ‘Senior Management Staff’ are responsible for the economic domain, the ‘Evaluation Section’ together with the ‘Ecology Staff’ are responsible for ecology and ethology, and the ‘Evaluation Section’ is responsible for social affairs. Contractors are typically the FOAG-assigned Research Stations or other Federal research institutes. For each evaluation, an advisory committee with four to five members is assembled. Its members consist of deputies from the Research Stations or from other relevant agencies such as the Federal Veterinary Office or the Federal Agency for the Environment, Forests and Landscape. An advisory committee usually meets at least once a year, and its duty is to:

- put the mandate to evaluate agricultural policy measures in concrete form;
- take decisions during the different evaluation phases;
- accept intermediate and final reports; and
- work together with the evaluation teams.

Established ways to disseminate evaluation results do not exist in the FOAG. The value given to these evaluations is highly dependent on the relative importance of the issue at hand to the FOAG or the general public as well as the degree of engagement of the project leader. Evaluation results are taken into consideration in internal annual planning, and when ordinances and laws are being elaborated or revised, evaluation results can eventually be inserted. Evaluation reports are sent to the departmental ‘General Secretariat’ as well. FOAG evaluations are open to the public, and they often serve as basic information in the relationship between the FOAG and policy beneficiaries, e.g. farmers or farm associations. In the ecological domain, evaluations serve as control instruments and as legitimization for adopted agricultural policy.

The aforementioned considerations point to a decentralized institutional evaluation model within the FOAG. The main purpose of evaluations is accountability. Results and recommendations are utilized in an instrumental and legitimizing way. When the purpose of evaluations is improvement, evaluation findings are also utilized in an interactive way, for example when the empirical measurement of policy aims is discussed among stakeholders and other interested parties.

Federal Office for Energy

Between 1990 and 2000, evaluation in the Federal Office for Energy (FOE) took place exclusively within the scope of the action programme ‘Energy 2000’. The FOE commissioned 58 evaluation studies, at a cost of SFr. 6.3 million, an amount that corresponded to 1.1 percent of the action programme’s total budget. External contractors conducted all of the ‘Energy 2000’ evaluation studies. FOE officials claim that internal evaluations would not have raised appropriate questions and would have been too uncritical.

A very important role in the FOE’s evaluation setting was assigned to the head of the action programme Energy 2000, who was the deputy director of the FOE. He was responsible for the execution of the evaluations and received

recommendations from an advisory committee consisting of the head of the action programme, two FOE officials, three independent experts, and an external evaluation consultant. The advisory committee was responsible for setting up the annual evaluation programme and for budget decisions about individual evaluations. The external evaluation consultant accompanied the different evaluations and was responsible for ensuring adherence to scientific standards, as well as being involved with setting up the annual evaluation programme and taking part in selecting evaluation proposals. From the beginning of 'Energy 2000', a position for evaluation was created within the FOE. This official was responsible for making data available to the evaluators, and both wrote contracts and helped select evaluation teams.

The external evaluation consultant in charge of 'Energy 2000' summarizes the most important effects of the evaluation activities within the programme as follows (Balthasar, 2000; cf. Balthasar and Rieder, 2000).

- First, evaluations had a positive influence on programme planning (evaluation as an information and control tool). In some cases, evaluation supplied basic information for necessary programme adaptations (e.g. reorganization of public energy information).
- Second, systematic evaluation had a positive influence on the formulation of the new energy law.
- Third, evaluations provided a systematic overview of the effects of measures and actions taken in energy policy and served as legitimation with regard to the electorate, parliament and the executive.
- Fourth, evaluation conveyed a positive image about Swiss energy policy abroad.

Further to these positive effects of the 'Energy 2000' evaluations, the consultant also underscored some of their weaknesses. Methodologically, most of the 58 evaluations attained a high quality standard. Nevertheless, recommendations were sometimes hard to understand and difficult to implement. Several evaluations were judged as too 'scientific' and too removed from practical needs. The independence of the evaluation was not always guaranteed, and in some cases, officials in charge of programmes tried to influence the conclusions. In specific cases, evaluations were delayed. Implementation of evaluation results depended on the readiness of the actors in charge. Sometimes those who were involved felt criticized by evaluations and that had the effect of hindering the utilization process.

The fact that one of the most important actors in the FOE's evaluation arrangement was the office's deputy director points to a centralized evaluation setting. The main purposes of 'Energy 2000' evaluations have been accountability and improvement. Evaluations have been mainly utilized in an instrumental way, and to a lesser extent in an interactive manner (cf. Balthasar and Rieder, 2000).

Comparison of Findings

Integration of Evaluation into the Political Decision-making Process

Evaluation activities in the federal administration are very varied. There is no consensus about what evaluation is or should do. Evaluation emerged for different reasons in the various administrative units, which has led to considerable variety in the functions evaluations are meant to perform. This assessment is particularly relevant for understanding how evaluations are institutionally embedded. One aspect is of course related to the respective availability of evaluation-related resources, which include technical expertise, funding, organizational know-how and legal advice, in the various federal offices. Nevertheless, in the last decade the Swiss evaluation community has developed a remarkable degree of professionalization – not only with regard to the execution of evaluations but also with respect to evaluation know-how within the federal administration. Yet evaluation findings and recommendations are used only rarely in the decision-making process, and when they are used, it is then often as a means of legitimation. Federal evaluation activities are strongly related to ‘power games’: evaluation often is seen as a way to gain power and influence within the administration. Because of this, evaluation co-ordination has to be considered as very weak. Prospective planning of evaluation activities synchronized with other administrative units does not usually occur.

Purpose, Institutional Settings and Utilization – Experiences Reconsidered

Following the empirical investigation of the evaluation practices with respect to purpose, utilization and institutional setting, we now compare these findings with the framework presented above. The overview highlights the points that follow.

1. Of the four distinct types of purpose in our analysis, only two are relevant. It was assumed that accountability and improvement are the most important (and legitimated) in an applied setting. On the other hand, the use of the instrument of evaluation for strategic purposes is hard to observe from without and – for those involved within the administration – hard to confirm. Furthermore, it is reasonable to assume that acquiring basic knowledge is not the highest priority in an applied setting.
2. The instrumental use of evaluations is the most common, though some offices seem to use evaluations in an interactive or legitimizing fashion. In contrast and as expected, conceptual as well as tactical utilization seemingly does not occur at all. This observation needs to be put in perspective, however, since there is a similarity between the strategic purpose and the tactical utilization of evaluations. Both aspects are judged as a less legitimate way of employing evaluation. Conceptual utilization is also problematic since little empirical evidence exists for the conceptual use of evaluations.
3. The two distinct institutional settings – centralized and decentralized – seem to be of no relevance with respect to evaluation purpose and utilization. We observed a single example in which a combination of accountability purpose

Table 2. Purpose, Institutional Setting and Utilization in Eight Federal Offices

Purpose	Utilization				
	Instrumental	Conceptual	Interactive	Legitimizing	Tactical
Accountability	C: SDC, FFA, FOE, FOJ D: FOAG		D: FOAG	C: FSIO, Seco D: FOPH, FOAG	
Improvement	C: FSIO, Seco, FOE D: SDC, FOPH		C: FOE D: SDC		
Basic knowledge					
Strategy					

Notes:

C: centralized institutional setting; D: decentralized institutional setting; FFA: Federal Finance Administration; FOAG: Federal Office of Agriculture; FOE: Federal Office for Energy; FOJ: Federal Office of Justice; FOPH: Federal Office of Public Health; FSIO: Federal Social Insurance Office; SDC: Swiss Agency for Development and Cooperation; Seco: State Secretariat for Economic Affairs

and interactive utilization was associated with only one of the settings, i.e. decentralized. In all other cases, one office worked within a centralized setting while another office was in a decentralized setting.

4. The most striking aspect of this compilation, in our view, is the fact that combinations of purpose and utilization exist that were unexpected. It is surprising, for example, that there are cases in which the purpose of accountability is combined with an instrumental utilization or even an interactive utilization.

The third of these observations is, in our opinion, of greatest theoretical as well as practical relevance. Bookcases full of evaluation literature have discussed the questions of purpose and especially utilization, but if there is a common theme in this literature, it is that it is important to have an appropriate setting for a given evaluation purpose or use (e.g. Patton, 1997; Sonnichsen, 1999: 60; Weiss, 1998; Widmer, 2002; Williams et al., 2002: 43–6).

There is no doubt that a lack of fit between the chosen institutional settings and the given purpose and proposed utilization reduces the possibility of the optimal use of evaluation.

Conclusions

Evaluation is a tool that can be applied for different purposes and with distinct uses. Our inquiry into different federal agencies in the Swiss Confederation has shown that this functional differentiation of evaluation is to a limited extent mirrored in the way institutional settings for evaluation are defined. Furthermore,

Evaluation 10(4)

these distinctions are only partly recognized by the officials. The belief that a single evaluation is able to fulfil all different purposes and forms of utilization at once is not uncommon (cf. *inter alia* Feinstein, 2002: 433–4), although the different functions compete strongly with one another. An improved functional differentiation makes it possible to reduce the observed weaknesses in federal evaluation practice. Two examples may make this clear:

- *Evaluation quality*: standards of evaluation quality depend on the functions attributed to a specific evaluation. Independence is of central importance for evaluations that are meant to ensure accountability. Closer identification and empathy is of central importance for evaluations that aim at improvement.
- *User groups and evaluation use*: user groups have a distinct importance depending upon the function of evaluation, which determines to a great extent the beneficiaries of evaluation. The function is also of relevance in terms of the way in which evaluation results and recommendations are used. In a decentralized setting, it is much easier for evaluators to establish a good relationship with the people involved, and this is one of the cornerstones in establishing a learning-oriented evaluation culture. On the other hand, a close relationship between the evaluation and those in charge of the evaluandum is a clear disadvantage when aiming for trust in evaluation results among external users in an accountability-oriented evaluation.

The purpose and utilization of an evaluation should determine the way in which evaluation is embedded institutionally. In practice, the three dimensions – purpose, institutional design and utilization – are often combined in an unsystematic fashion. Our argument is that the purpose and the (intended) utilization should be the starting point in defining the most appropriate institutional design. The responsible units should take into account the purpose and utilization of an evaluation. Only once these considerations have been addressed should they decide upon the most suitable institutional setting for the evaluation. In this sense, there is a strong need for functional differentiation to provide the conditions under which evaluation is able to fulfil the expectations put forward by various stakeholders.

In addition the decision makers should acknowledge that defining the purpose as well as the intended utilization of an evaluation is essential. The ‘one-structure-fits-all’, fully integrated, multi-purpose and overall utilization evaluation is an illusion that hinders the development of good-quality evaluation.

The empirical investigation of the evaluation practices within the Swiss Federation showed considerable diversity. The empirical materials allowed us to investigate a question of general interest, namely how the relationship between the purpose and the utilization of evaluation is related to institutional and organizational design. Although empirical research was focused on a specific national case, we believe its findings are relevant not only for decision makers within the Swiss Federation but also for the broader evaluation community.

Note

1. Findings presented in this article are based on a study commissioned by the Swiss Federal Chancellery. The results of the study were published in a more detailed report (see Widmer et al., 2001) that also describes the evaluation activities of the Federal Audit Office and the Parliamentary Administration Control (cf. Janett, 2000) and discusses other topics, including evaluation quality (see Widmer, 1996 and forthcoming). The authors would like to thank Erwin Rüegg and John Bendix for their support as well as the three anonymous reviewers for their helpful comments.

References

- Balthasar, A. (2000) *Energie 2000*. Zurich: Rüeegger.
- Balthasar, A. and S. Rieder (2000) 'Learning from Evaluations. Effects of the Evaluation of the Swiss Energy 2000 Programme', *Evaluation* 6(3): 245–60.
- Bussmann, W. (1995) 'Evaluation and Grassroots Politics: The Case of Switzerland', *Knowledge and Policy* 8(3): 85–98.
- Bussmann, W. (1996) 'Democracy and Evaluation's Contribution to Negotiation, Empowerment and Information', *Evaluation* 2(3): 307–19.
- Derlien, H.-U. (1997) 'Die Entwicklung von Evaluationen im internationalen Kontext', in W. Bussmann, U. Klöti and P. Knoepfel (eds) *Einführung in die Politikevaluation*, pp. 4–12. Basel: Helbling & Lichtenhahn.
- Eggers, H. W. (2002) 'Project Cycle Management', *Evaluation* 8(4): 496–504.
- Feinstein, O. N. (2002) 'Use of Evaluations and the Evaluation of their Use', *Evaluation* 8(4): 433–9.
- Fornerod, S. (2001) *A Quoi et a Qui Servent les Evaluations?* Cahier de l'IDHEAP no. 199. Chavannes-près-Renens: IDHEAP.
- Furubo, J.-E., R. Rist and R. Sandahl, eds (2002) *International Atlas of Evaluation*. New Brunswick, NJ: Transaction.
- Horber-Papazian, K. and L. Thévoz (1990) 'Switzerland: Moving Towards Evaluation', in R. C. Rist (ed.) *Program Evaluation and the Management of Government*, pp. 133–43. New Brunswick, NJ: Transaction.
- Janett, D. (2000) 'Parliamentary Evaluation in Switzerland and its Effects on Legislative and Administrative Action', Paper presented at the European Evaluation Society Conference, Lausanne: 12–14 October.
- Leeuw, F. L. (2000) 'Evaluation in Europe', in R. Stockmann (ed.) *Evaluationsforschung*, pp. 57–76. Opladen: Leske and Budrich.
- Linder, W. (1994) *Swiss Democracy*. New York: St Martin's Press.
- Mayne, J. (1994) 'Utilizing Evaluation in Organizations', in F. L. Leeuw, R. Rist and R. C. Sonnichsen (eds) *Can Governments Learn?* pp. 17–43. New Brunswick, NJ: Transaction.
- OECD's Development Assistance Committee (2000) 'Development Co-operation Review of Switzerland', *DAC Journal* 1(4) (see www.oecd.org).
- Owen, J. M. and P. J. Rogers (1999) *Program Evaluation*. London: Sage.
- Patton, M. Q. (1997) *Utilization-focused Evaluation*. Thousand Oaks, CA: Sage.
- Scriven, M. (1991) *Evaluation Thesaurus*, 4th edn. Newbury Park, CA: Sage.
- Sonnichsen, R. C. (1999) 'Building Evaluation Capacity Within Organizations', in R. Boyle and D. Lemaire (eds) *Building Effective Evaluation Capacity*, pp. 53–73. New Brunswick, NJ: Transaction.

- Spinatsch, M. (2002) 'Evaluation in Switzerland: Moving Toward a Decentralized System', in J.-E. Furubo, R. Rist and R. Sandahl (eds) *International Atlas of Evaluation*, pp. 375–91. New Brunswick, NJ: Transaction.
- Vedung, E. (1997) *Public Policy and Program Evaluation*. New Brunswick, NJ: Transaction.
- Weiss, C. H. (1977) 'Research for Policy's Sake: The Enlightenment Function of Social Science Research', *Policy Analysis* 3(4): 531–45.
- Weiss, C. H. (1980) 'Knowledge Creep and Decision Accretion', *Knowledge* 1(3): 381–404.
- Weiss, C. H. (1981) 'Use of Social Science Research in Organizations, the Constrained Repertoire Theory', in H. D. Stein (ed.) *Organization and the Human Services*. Philadelphia, PA: Temple University Press.
- Weiss, C. H. (1988) 'If Program Decisions Hinged Only on Information', in M. C. Alkin (ed.) *Debate on Evaluation*, pp. 208–22. Newbury Park, CA: Sage.
- Weiss, C. H. (1989) 'Congressional Committees as Users of Analysis', *Journal of Policy Analysis and Management* 8(3): 411–31.
- Weiss, C. H. (1995) 'The Haphazard Connection: Social Science and Public Policy', *International Journal of Educational Research* 23(2): 137–50.
- Weiss, C. H. (1998) *Evaluation*, 2nd edn. Upper Saddle River, NJ: Prentice-Hall.
- Widmer, T. (1996) *Meta-Evaluation*. Bern: Haupt.
- Widmer, T. (2000) 'Evaluating Evaluations: Does the Swiss Practice Live up to the "Program Evaluation Standards"?', in C. Russon (ed.) *The Program Evaluation Standards in International Settings*, pp. 67–80. Kalamazoo, MI: Western Michigan University, Evaluation Center.
- Widmer, T. (2002) 'Staatsreformen und Evaluation', *Zeitschrift für Evaluation* 2002(1): 101–14.
- Widmer, T. (forthcoming) 'Instruments and Procedures for Assuring Evaluation Quality', in R. Schwartz, J. Mayne and J. Toulemonde (eds) *Assuring the Quality of Evaluative Information: Prospects and Pitfalls*. New Brunswick, NJ: Transaction.
- Widmer, T., C. Landert and N. Bachmann (2000) *Evaluation Standards of SEVAL, the Swiss Evaluation Society (SEVAL Standards)*. Bern/Geneva: SEVAL. Available at: http://www.seval.ch/en/documents/SEVAL_Standards_2000_en.pdf (site visited: 25/09/04).
- Widmer, T., E. Rüegg and P. Neuenschwander (2001) *Stand und Aussichten der Evaluation beim Bund*. Bern: Bundeskanzlei.
- Williams, K., B. de Laat and E. Stern (2002) *The Use of Evaluation in the Commission Services*. Brussels: European Commission, DG Budget.
- Zimmermann, W., D. Kolly and P. Mahon (1989) *Die wissenschaftliche Evaluation von Bundespolitik*. Bern: EDMZ.

THOMAS WIDMER is a senior research fellow, a lecturer, and the head of the Research Area 'Policy Analysis and Evaluation' at the Department of Political Science, University of Zurich, and Affiliate, Harvard University, Cambridge MA. He is responsible for various evaluation studies and serves as a consultant to several public bodies in evaluation questions; he also serves on the boards of the Swiss and the European Evaluation Societies. Please send correspondence to: Department of Political Science, University of Zurich, Seilergraben 53, CH-8001 Zurich, Switzerland. [email: thow@pwi.unizh.ch]

Widmer and Neuenschwander: Embedding Evaluation in the Swiss Administration

PETER NEUENSCHWANDER is a research fellow in the Research Area 'Policy Analysis and Evaluation', Department of Political Science, University of Zurich. He has 10 years' experience in policy evaluation in Switzerland, with an emphasis on the environment as well as on HIV/AIDS and drug policy. Please send correspondence to: Department of Political Science, University of Zurich, Seilergraben 53, CH-8001 Zurich, Switzerland. [email: peneu@pwi.unizh.ch]