

The Effects of the Institutional Design on the Utilization of Evaluation: Evidenced Using Qualitative Comparative Analysis (QCA)

The International Journal of Theory, Research and Practice, 12(3), 2006, London.

Andreas Balthasar, Interface Institute for Policy Studies, Switzerland

Abstract

This paper aims to present some of the results from a study in progress, focusing on the influence of the distance between evaluators and evaluatees on the utilization of evaluations. The basis for the results presented here is an analysis of ten case studies from Switzerland. These cases involve evaluations that were assigned in different institutional contexts, with widely varying distances between evaluators and evaluatees. "Qualitative Comparative Analysis" (QCA) has been used to interpret the cases, in order to allow a combination of case- and variable-centred comparisons. The analysis indicates that under certain conditions, the distance between evaluators and evaluatees has no influence on the use of evaluations. In particular, formative objectives can be achieved quite independently of the distance. When interpreting the results, however, one should not neglect the fact that they are solely based on a systematic evaluation of ten case studies with QCA. Generalization is not permissible on this basis, nor is this the aim of the present article. On the contrary, the objective is to continue developing the debate about the influence of the distance between evaluators and evaluatees on the utilization of evaluations.

Keywords: evaluation utilization, qualitative comparative analysis (QCA), institutional design, Swiss federal administration, Switzerland

1. Introduction and question to be answered¹

Whenever evaluations are arranged and carried out, care is taken to ensure that they yield results which can be utilized for policy purposes. The utilization of evaluations is one of the most prominent issues in the international literature on evaluation. While there is extensive literature on types of evaluation usage (Widmer et al. 2004; Weiss et al. 2005) and the mechanisms governing the outcomes of evaluation (Henry et al. 2003), only a few contributions can be found dealing with the influence of evaluation-specific institutional design on utilization (Rogers et al. 1995; Williams et al. 2002). However, this design normally defines the distance between evaluators and evaluatees. This distance is a highly important factor of influence on the utilization of evaluations. In a recent article in this journal, Widmer and Neuenschwander (2004) presented some insights into the relationship between utilization and institutional design in analyzing Federal Swiss evaluation practice. They conclude that the lack of use of an evaluation is often due to a bad match between the institutional setting and the given purpose of an evaluation. The present article develops this issue further, focusing on the influence exerted by the distance between the evaluator and the evaluatee on the intensity of the utilization of evaluations. This analysis is presented with the help of "Qualitative Comparative Analysis" (QCA), a method that allows a combination of case- and variable-centred comparisons.

We present our analytical framework in the first sections of the article, followed by descriptions of the method and database. After presenting the results of the analysis, we conclude with a synthesis of the findings and a discussion of the method.

2. Analytical framework

The utilization of evaluations can be investigated in very different ways. Recent contributions by Henry and Mark have advanced this discussion, proposing a model that will lead to greater systematization of research on the subject (Henry 2000; Henry et al. 2003; Mark et al. 2004). This work recombines some of the existing categories of evaluation use, and it attempts to make the processes by which influence is exerted both conceptually distinct and better integrated (Weiss et al. 2005: 14). However, a simpler approach seemed adequate for the purpose of our analysis. This approach distinguishes four types of evaluation usage (instrumental, conceptual, process and symbolic use) and three categories of factors that influence the use of evaluation results.

2.1 Types of evaluation usage

Much of the experience described in the empirical literature, together with a great deal of practical experience, is captured by the *four types of evaluation usage* which we use. This classification is very similar to a distinction proposed by Vedung (Vedung 1997; Vedung 1999) which was recently used by Widmer and Neuenschwander (Widmer et al. 2004), and also by Weiss (Weiss et al. 2005):

- the first type is *instrumental use* for policy decision-making and problem-solving. Instrumental use is defined as "using evaluation results as direction for decision making" (Weiss et al. 2005: 13). Accordingly, the term "instrumental use" may be applied if evaluations provide impetus that is intentionally incorporated into the policy formulation processes in a manner which can be proven, for example by the implementation of their recommendations.
- the second type of use is described as the *conceptual use* of results: Weiss (1977) defines this type of use as "gradual sedimentation of insights, theories, concepts and ways of looking at the world". Weiss has made particular reference to this form of using evaluation results, and she has coined the term "enlightenment" for it.
- the third form is the *symbolic use* of evaluation results. This term may be used when decision-makers use evaluations in order to confirm their perspective and to obtain legitimation for themselves (Weiss et al. 2005: 13). With reference to Owen (Owen 1993), Clarke (Clarke 1999: 177) mentions the "'symbolic use' of evaluation research, [in] the case when an evaluation is commissioned by an individual or organization for reasons other than the improvement of a program or service".
- Fourth, authors such as Patton (Patton 1997; Patton 1998) and Wholey (Wholey 1994) point out that the *process-related use* of evaluation must not be underestimated. Forss and colleagues distinguish five types of process use: "learning to learn; developing networks; creating shared understanding; strengthening the project; and boosting morale" (Forss et al. 2002: 29). In the view of Weiss et al. (2005: 14), process use is not parallel to the other three forms because it is a result of a process, not of an evaluation.

2.2 Factors of influence on the use of evaluations

Research has also devoted intensive attention to numerous *factors of influence on the use of evaluations*. Alkin (Alkin et al. 1985), for example, differentiates 50 factors that help determine the use of evaluation. Huberman and Gather Thurler (Huberman et al. 1991) identify 47 factors of influence that have proven significant in terms of implementation. Shulha and Cousins (Shulha et al. 1997: 196) have summarized the results of scientific research into the use of evaluations. They compiled a sort of "shopping list" of factors that influence the use of evaluation results. On the basis of this and other literature (e.g. Henry et al. 2003), we differentiated *three categories of factors that influence the use of evaluation results*: the institutional context, the evaluation environment and the evaluation process. This is obviously a massive simplification of the reality, but it was necessary in order to permit a systematic comparison of the chosen cases.

- *Institutional context*: every public administrative institution (e.g. a Federal agency or a monitoring body) has a more or less formalized practice – regardless of the specific evaluation project – that specifies how evaluations are to be prepared and carried out. This practice, which regulates the distance between the evaluator and the evaluatee, may be described as the institutional context of an evaluation.

- *The environment of the evaluation*: there is no doubt that the environment of a specific evaluation also influences its use. In this connection, Kingdon mentions the importance of "windows of opportunity" and stresses that new scientific arguments can only be used in politics if suitable opportunities arise to do so (Kingdon 1984). Lester and Wilds (Lester et al. 1990: 315) speak of "contextual variables": "These factors relate to the nature of the political environment within which policy analysis occurs: nature of the problem, politically feasible, immediate decision needed, centralized decision-making, single agency issue, proposed funding change, amount of conflict, issue salience and bureaucratic variables, user characteristics, clear definition of objectives by decision-maker, decision-maker interest, decision-maker style, decision-maker participation."

- *The process of evaluation*: various studies also suggest that factors relating to the manner of undertaking an evaluation are very important as regards the use of the evaluation results. These include (for example) the works by Greenberg and Mandell (Greenberg et al. 1995). Lester and Wilds (Lester et al. 1990) call these influences "technical variables". These variables relate to the quality of an evaluation as defined by the "Joint Committee on Standards for Educational Evaluation" (Joint Committee on Standards for Educational Evaluation et al. 2000). These standards are based on a fundamental assumption: in order to be considered as of high quality, the process of an evaluation must simultaneously satisfy requirements regarding usefulness, implementability, correctness and accuracy (cf. Widmer 2000).

These factors were taken as the basis for ten case studies dealing with evaluation practice in the Swiss Federal Administration. The next section describes how these cases were selected and analyzed.

3. Method

3.1 Database

A three-stage process was used to select the case studies:

In the first stage, we assembled a comprehensive list of all the evaluations undertaken in the Federal Administration between 1999 and 2002. This list comprises about 340 studies. 32 of the total of 46 Swiss Federal agencies carried out evaluations during the period examined (cf. Balthasar 2003);

In the second stage, we selected ten of the 32 Federal authorities who had carried out at least one evaluation. The following two criteria were taken into account for the selection: firstly, the Federal authorities had to differ in terms of their institutional context, in order to ensure maximum variability of the influencing variables that interest us here. In this phase of the project, the information used to assess the institutional context was based on appraisals obtained from discussions with experts and from the literature (e.g. Widmer et al. 2001). Secondly, we took account of the requirement that the Confederation's control bodies and all seven of its Departments had to be represented in the case studies. The following box gives a brief presentation of the ten institutions we chose for further investigation, indicating the distance between evaluators and evaluatees in each context. This indication is based on expert interviews and an assessments of the author about the institutional context key influencing factors described in section 3.2:

Box 1: Institutional settings under investigation

Institution	Short description of the institution	Distance
Parliamentary Control of the Administration (PCA)	The PCA is the competence centre of the Federal Assembly responsible for evaluations. It supports parliamentary supervision through scientific assessments and it evaluates the concepts, implementation and impact of the measures taken by the Federal authorities. The evaluation subjects are defined by the Control Committees (CC) of the Federal Assembly.	large
Swiss Federal Audit Office (SFAO)	The SFAO has a Competence Centre for Performance Audit and Evaluation to scrutinize the completeness and impact of Confederation initiatives with particular financial significance. In this way, the SFAO takes on the duties of an autonomous court of auditors in Switzerland. The selection of subjects for evaluation by the SFAO and the financing of the studies are handled on an autonomous basis.	large
Swiss Agency for Development and Cooperation (SDC)	The SDC has been dealing with evaluation for about thirty years. Nowadays, evaluation functions are embedded in Program Cycle Management (Lehmann et al. 2004). The SDC differentiates between different forms of evaluation. In most cases, the responsibility for carrying out an evaluation rests with the sections that are responsible for managing the programs under examination. Financing is provided from the budget for the program.	small
Swiss Federal Office of Public Health (SFOPH)	Evaluation began to play a part at the SFOPH in the late 1980s (Widmer et al. 2004). Today the directorate-level "Centre de Compétences en Evaluation" supports all the SFOPH's evaluation activities, acting as an intermediary between practitioners and scientists in order to define the objectives and focus of a study. Nevertheless, the administrative units are still responsible for clarifying the requirements for evaluations, and for their content. It is also up to them to finance evaluations.	quite large
Federal Office for Migration (FOM)	Evaluations have not played a very important part in the FOM up to now, and they are carried out infrequently. The FOM still does not have a central contact person or a budget for evaluations. Staff with responsibility for the actual measures are also responsible for planning and implementing any evaluations.	small
Inspectorate of the Federal Department of Defence, Civil Protection and Sports (DDPS)	The Inspectorate of the DDPS is the Department's internal audit body. The Inspectorate has its own budget. The audit subjects are specified by the Head of the Department - a fact which can reduce the autonomy of the inspectorate. Apart from this, the Inspectorate is independent from the rest of the Department.	quite large
Swiss Federal Office of Personnel (FOP)	The FOP has neither a central unit with responsibility for evaluations nor a budget for this purpose. Evaluations consist either of specific assessments of representative personnel surveys carried out by the FOP, or of works which the FOP has compiled on behalf of the Swiss Federal Council. The FOP carries out most evaluations itself	quite small

	and has no special budget for this activity.	
Swiss Federal Veterinary Office (SFVO)	The SFVO is mainly concerned with the health of animals in Switzerland. This brief requires numerous monitoring activities to ensure ongoing supervision – for example, in connection with combating epidemics. Ex-post policy evaluations as such are rare. The SFVO does not have a central unit with responsibility for evaluations. The points of contact are the individuals responsible for the Controlling and Research sections; they cooperate very closely with the staff who are responsible for the measures.	quite small
Federal Office of Transport (FOT)	Until recently, no systematic evaluations were carried out in the FOT. This means that the Office does not have a culture of ascertaining the effects of measures after they have been implemented. Until 2004, the Office had neither a central unit responsible for evaluation nor a dedicated budget for evaluations.	small
Swiss Federal Office of Energy (SFOE)	Systematic evaluation in the Swiss Federal Office of Energy started in 1991 (Balthasar 2000: 201-220). The senior evaluation authority is the Head of the Energy Policy Department, who makes the final decisions on the implementation of studies. The SFOE has a unit responsible for evaluation, which works as an interface between the internal Office players and the evaluators. The SFOE has a special budget dedicated to evaluation.	quite large

In the third stage, a random process was used to identify one evaluation for each of the selected authorities, to be examined in an in-depth case study. The random selection process was chosen in this stage of the election process as this was considered to be the best way of finding a typical case for each of the ten selected institutional settings. The following box briefly presents the cases that were studied. The aim here is not to deal systematically with all the variables, but to give an impression of the variety of evaluations. A separate column indicates the general assessment of the utility of each evaluation. This assessment was made by the agent responsible for managing the evaluation (see section 3.2)

Box 2: Evaluations investigated

Institution	Short description of the evaluation	Assessment of utility
Parliamentary Control of the Administration (PCA)	In 2000, the Control Committee (CC) of the Swiss National Council commissioned the PCA to compile a report on the implementation of Swiss cartel law (Parlamentarische Verwaltungskontrolle 2000). This report identified numerous problems regarding the implementation of Swiss cartel law. The level of use of this study was assessed as rather low. The reason was that the parliamentary discussion process had already led to similar assessments before the study was completed, so this evaluation came too late to reveal anything new.	rather low
Swiss Federal Audit Office (SFAO)	In 2001, the SFAO carried out an analysis of the Confederation's subsidies for permanent workshops for disabled people. These workshops – about 300 in number – offer disabled persons the chance to carry out useful activities that provide them with an income. The analysis by the SFOA did not reveal any fundamental errors, but it did identify some weak points and substantial possibilities for improvement (Eidgenössische Finanzkontrolle 2002). The main problem regarding the use of this report was that responsibility for the permanent workshops had to be shifted from the Confederation to the Cantons as part of a redistribution of tasks.	rather low
Swiss Agency for Development and Cooperation (SDC)	The evaluation under investigation must remain anonymous since it ended in a conflict between the SDC and the evaluated private partner. This evaluation assesses the results delivered by this partner, the partner's entrepreneurial capacity and competitiveness, and the relevance and significance of the different mandates, with recommendations regarding future collaboration. The evaluation was undertaken by a team that was commissioned without a formal invitation to tender.	high
Swiss Federal Office of Public Health (SFOPH)	The evaluation of the 2001 National Flu Prevention Campaign was triggered by the "Centre de Compétences en Evaluation" of the SFOPH. The evaluation was carried out by specialists external to the administration who were selected in a competitive procedure. One aim of the study was to supply information about doctors' acceptance of the campaign and its messages (Sardi et al. 2002). Most of the suggested recommendations were implemented.	quite high
Federal Office for Migration (FOM)	The evaluation of counselling on returning home in the asylum sector was triggered by the section responsible for this subject area. The study was carried out externally to the administration (Jaggi et al. 2000). The evaluators were designated without obtaining different offers. A group of responsible FOM staff provided intensive follow-through for	quite high

	the study. The evaluation singled out some possible ways of optimizing the way this assistance is provided, and the majority of the recommendations were implemented.	
Inspectorate of the Federal Department of Defence, Civil Protection and Sports (DDPS)	As a basic rule, the internal audit reports of the DDPS are not accessible to third parties, except members of the Federal Parliament. The report whose use we examined in depth took a very critical look at the activities which it assessed. Nevertheless, it was well received by most of the parties involved, and most of the recommendations it proposed were implemented.	quite high
Swiss Federal Office of Personnel (FOP)	The case study deals with the evaluation of the effectiveness of measures to promote women in the Federal Administration. The evaluation was carried out internally, mainly on the basis of data from the central personnel administration and a representative written survey in four offices (Eidgenössisches Personalamt 2002). A group of personnel managers provided intensive follow-through for the study. It contains various recommendations, but only a small proportion of them were implemented.	quite low
Swiss Federal Veterinary Office (SFVO)	A comparatively new methodological approach was applied in the evaluation of a measure to promote the keeping of milk-cows in a manner suitable to the species (Danuser et al. 2002). Follow-through and financing were jointly provided by the SFVO and the Swiss Federal Office for Agriculture. Most of the evaluation was undertaken by a doctoral student who was writing a dissertation and working in the SFVO. No competitive procedure to select the evaluation team was carried out. In this study, practice-oriented recommendations are given lower priority than the scientific presentation of the results.	quite high
Federal Office of Transport (FOT)	The case study concerns an evaluation of the effects of measures implemented to shift goods traffic from railways to roads. The report was compiled by the FOT itself, within the administration, for the Swiss Federal Council and the Parliament. The report contains a number of recommendations, most of which were implemented (Schweizerischer Bundesrat 2002).	high
Swiss Federal Office of Energy (SFOE)	In 1998, the SFOE commissioned an evaluation of the strategy and services used to implement energy-saving measures in small and medium-sized enterprises (Gerheuser 1999). The work had to be carried out against a background of mistrust and lack of interest from the staff responsible for this initiative. The report arrives at a critical overall assessment of the activities supported by the Federal Office. This evaluation had some influence at the conceptual level but in general terms, its use is assessed as rather low.	rather low

3.2 Definition of factors and data collection

It is clear that the case studies yielded insights into a great variety of outcomes from the analyzed evaluations, with varying degrees of intensity. The studies also highlight a large number of influencing factors. To arrive at a systematic and transparent comparison of the ten cases, it was necessary to identify factors and categories on the basis of the analytical framework described in section 2.

In order to answer the question that was set, five target factors were defined together with four influencing factors originating from the institutional context, the evaluation environment and the evaluation process:

- *target factors relating to the types of use of evaluations*: the four types of use mentioned in Section 2 – namely "instrumental use" (INST), "conceptual use" (CONC), "process-related use" (PROC) and "symbolic use" (SYMB), together with a factor related to general use (GENE) – were defined as target factors. Interest focuses on the combinations of the factors on which the type and intensity of use depend. For each of the five factors, two levels of intensity were defined (high or low intensity of use). This simplification was necessary to overcome a restriction inherent in the QCA method which we intended to apply.
- *the "institutional context" key influencing factor*: in order to identify the relevant factors that can define the distance between evaluators and evaluatees, we analyzed the available literature (Rogers et al. 1995; Widmer et al. 2004) and conducted discussions with experts. This preliminary work led us to the following six factors:

- *the existence of a unit responsible for the evaluation*: we assumed that the distance between evaluators and evaluatees is greater if a unit located outside of line responsibility is in charge of administering the project than if the program director handles the evaluation himself.
- *the existence of an evaluation budget*: we further assumed that the distance between evaluators and evaluatees is greater if the evaluations do not have to be financed from the funds for the evaluated action.
- *way of triggering the evaluation*: we assumed that in a case where an evaluation unit or a decision-maker can trigger the evaluations, the distance between evaluators and evaluatees is greater than in a case where the individuals responsible for the programs or actions can trigger it.
- *allocation practice*: it is assumed that the distance between evaluators and evaluatees is greater if there is a competitive procedure (invitation to tender) for assigning evaluations than in a case where projects are directly awarded to evaluators who appear suitable.
- *publication practice*: we further assumed that the distance between evaluators and evaluatees is greater if the reports are published systematically. In this case, the evaluator not only has a responsibility to the contractor but also to the public and the evaluation community who will judge the quality of its work.
- *role of external evaluation*: the applicable assumption is that the distance between evaluators and evaluatees is greater for external evaluations. It is clear that if external evaluation is controlled by a strong internal steering group, it does not necessarily produce a large distance between evaluators and evaluatees. However, we assume that in any case, the distance when an external evaluator is present is greater than for a self-evaluation.

On the basis of these six factors, we formed the key influencing factor of "distance between the evaluator and the evaluatee" (DIST). This value expresses whether - and how - the evaluatees are able to influence the evaluation. We assume that the evaluatee's influence on the evaluation increases as the distance between the evaluator and the evaluatee reduces. The DIST factor has two levels of intensity: DIST is described as "large" if (in the case of the evaluation under consideration) at least four of the six questions about institutional factors described above can be answered in the affirmative. On the other hand, DIST is "small" if only one or two of these questions can be answered in the affirmative.

Influencing factors arising from the environment of the evaluation: as part of the present analysis, two factors arising from the environment of the evaluation are incorporated into the assessment: the differences between cases in terms of the purpose of the evaluation, and the question as to whether the evaluation was initiated in order to reach a decision in a specific context. These two factors were chosen from a total of six because they vary greatly from case to case:

- the "purpose of evaluation" factor (PURP) describes the basic objectives specified for an evaluation. In this case, a distinction is drawn between more summative (accountability-oriented) evaluations and more formative (optimizing) ones (Vedung 1997). In addition, some evaluations may have both a summative and a formative character. The "purpose of evaluation" factor therefore has three levels.
- the "routine" variable (ROUT) indicates whether or not the evaluation was carried out in a context where evaluation is part of the routine. Two levels are allowed (yes/more yes, more no/no).

- *influencing variables related to the evaluation process*: our analysis considers the "usefulness" (USEF) variable as representing the influence of the evaluation process on the use of evaluation results. "Usefulness" was chosen out of four factors describing the evaluation process because it varies greatly from case to case. This factor answers the question as to whether the evaluation is geared to the information requirements of its envisaged users and therefore was potential for implementation. For this reason, it incorporates the concerns of the usefulness standard, which aim to ensure that the evaluation is set up so that it can be useful (cf. Widmer 2000). The USEF variable includes two levels (yes / more yes, more no / no).

The data were collected in two stages:

- in the first stage, the listed variables for each of the ten selected evaluations were recorded during a semi-standardized interview conducted with the individual in the Federal authority who was responsible for managing the evaluation in question. It is clear that in many cases, this agent is not the most important user of an evaluation, but he is normally the person with the best knowledge about the institutional context of the case.
- in order to overcome the problem of the supposedly limited neutrality of the person in charge of a special evaluation, interviews were conducted with potential users of the evaluation in the second stage. Interestingly enough, there were some cases where the person in charge of the evaluation was also its best-informed user. In other cases, the person in charge of the evaluation was more critical about the use of the evaluation than the potential user who was interviewed.

3.3 Data analysis

The case studies were assessed using the "Qualitative Comparative Analysis" methodology. From a small number of cases, the QCA method makes it possible to derive factor combination patterns that link the cases (Ragin 1987; Varone et al. 2005; De Meur et al. 2002). QCA is based on Boolean algebra and is used to search systematically for commonalities and differences between cases. QCA indicates the importance of combinations of causes for specific results. The method starts out from the idea that observable effects depend on combination with other factors. Consequently, it does not involve the individual explanatory factors as such but rather the combinations in which these factors occur, which – according to the logic of QCA – offer the only possible way of explaining the particular result.ⁱⁱ We carried out the QCA analysis in three stages (De Meur et al. 2002; Ledermann 2004):

In the first stage, the configurations of factors for the ten cases are presented in a table with their respective levels of intensity. If two or more cases showed the same configuration, they were grouped together and shown as one configuration. *In the second stage of the work*, the cases observed in reality were minimized using an algorithm based on Boolean algebra. At this point, two configurations were combined to form one configuration if they only differed in respect of one level of a variable. As far as possible, this combination aims to minimize the number and complexity of the configurations for the respective level of each target outcome. *In the third stage of the analysis*, the factor combination patterns were simplified even further. In this step, those configurations which were theoretically possible but which were not observed in the ten cases (so-called "logical cases") were included in the minimization. In this context, it is assumed that the outcome of the combinations that were not observed contributes towards the simplification of the solution. If the assumptions do not contribute towards a simplification, it is supposed that the cases with the corresponding combination of independent factors (if any) would have the opposite outcome. In principle, the inclusion of logical

cases generalizes the explanatory patterns that are suggested by the observed cases. Although the method is no longer purely descriptive but inductive, the inclusion of logical configurations is nevertheless an instructive research step. At a general level, one could say that such a procedure is needed in any scientific enquiry, since scientific research *does* make inferences that go beyond the specific observations collected. Furthermore, using logical configurations can produce further research questions. However, one problem is that in some cases, simplifying solutions are only possible if contradictory logical cases are included. In this case, a careful interpretation of the results has to be combined with a re-examination of the real cases (Vanderborcht et al. 2003).

4. Results of the QCA analysis

The following analyses focus on the effects of the institutional context of an evaluation on the intensity of the various types of use of the evaluation results (instrumental, conceptual, process-related, symbolic, general), taking account of selected factors related to the evaluation environment and the evaluation process. Let us start by considering the configurations of variables for the ten cases that were studied.

Figure 2: Table showing influencing factors and target factors

Fall	Influencing factors				Target factors				
	DIST	PURP	ROUT	USEF	INST	CONC	PROC	SYMB	GENE
PCA	1	2	0	0	0	1	0	0	0
SFAO	1	2	1	1	0	1	0	0	0
SDC	0	0	1	0	1	1	0	0	1
SFOPH	1	0	1	1	1	0	1	0	1
FOM	0	0	0	1	1	1	1	0	1
DDPS	1	1	1	1	1	0	0	0	1
FOP	0	1	0	1	0	0	1	0	0
SFVO	0	1	1	1	0	0	1	0	1
FOT	0	2	0	1	1	1	1	1	1
SFOE	1	2	1	0	0	1	0	0	0

Figure 2 shows that ten different configurations of influencing variables with complete data records can be observed.

If we consider the instrumental use of the evaluation results (INST), five cases with a positive outcome can be identified. For the conceptual use of evaluations (CONC), six combinations with high intensity of use can be seen. Five cases show highly intensive process-oriented use of the evaluation results (PROC). As regards the fourth target variable - the symbolic use of evaluation results (SYMB) - there is one configuration which produced high intensity of use. On a general level, six evaluations were regarded as useful (GENE).

Now let us turn to the second and third stages of the analysis. For this purpose, we shall initially limit ourselves only to those cases that were observed in reality. Then we shall incorporate the "logical cases" into the analysis as well. All the solutions presented in this article were reached without any contradictory logical cases (cf. Section 3.3). To illustrate the method and the results, we are presenting the results for the "instrumental use" in detail. We shall merely summarize the results for the other elements of the analysis.

Results for "instrumental use"

The analysis of the five real cases with a positive outcome shows that the instrumental use of the results from the evaluations is *high*:

- if, *with a large distance* (DIST 1) between the evaluator and the evaluatee, the evaluation was carried out on a routine basis (ROUT 1), there was potential for implementation (USEF 1) and the evaluation was of either formative (PURP 0) or summative (PURP 1) character, *or*
- if, *with a small distance* (DIST 0) between the evaluator and the evaluatee, either the formative purpose (PURP 0) was prominent, the evaluation was carried out on a routine basis (ROUT 1) and there was no potential for implementation (USEF 0), or the evaluation pursued formative or formative and summative goals (PURP 0,2), was not carried out on a routine basis (ROUT 0) and had potential for implementation (USEF 1).

In formal terms, this result may be expressed as follows:

$$\text{INST}(1) = \text{DIST}(1) \text{ROUT}(1) \text{USEF}(1) [\text{PURP}(0) + \text{PURP}(1)] + \\ \text{DIST}(0) [\text{PURP}(0) \text{ROUT}(1) \text{USEF}(0) + \text{PURP}(0,2) \text{ROUT}(0) \text{USEF}(1)]$$

When 11 logical cases are included, the instrumental use is high:

- if the evaluation shows a formative aspect (PURP 0) *or* if the evaluation was carried out at a small institutional distance (DIST 0) and both formative and summative goals were pursued (PURP 2), *or* if the evaluation took place with a large institutional distance (DIST 1) and summative goals were pursued (PURP 1).

In formal terms, this result may be expressed as follows:

$$\text{INST}(1) = \text{PURP}(0) + \\ \text{DIST}(0) \text{PURP}(2) + \\ \text{DIST}(1) \text{PURP}(1)$$

In this way, with the help of QCA, the formative orientation of an evaluation can be identified as a sufficient condition for its instrumental use. This reflects the situation in the cases of the SDC, the SFOPH and the FOM, all of which primarily pursued formative goals and were used instrumentally. The result shows that the hypothesis asserting that the attainment of formative goals is reliant on a small distance between evaluator and evaluatee cannot be confirmed. A small distance (DIST 0) only led to high instrumental use when summative and formative goals were jointly pursued, as shown by the example of the FOT case. Hence, whereas formative goals led to instrumental use regardless of the distance between evaluator and evaluatee, the instrumental use of evaluations with summative objectives was reliant on a large distance between evaluator and evaluatee.

Now let us consider the five real cases with low instrumental use. The instrumental use is *low*:

- if, *with a large distance* between evaluator and evaluatee, the evaluation pursued summative and formative goals and either there was no potential for implementation or the evaluation took place on a routine basis, or

- if, *with a small distance* between evaluator and evaluatee, the evaluation pursued summative goals and there was potential for implementation.

In formal terms, this result may be expressed as follows:

$$\text{INST}(0) = \text{DIST}(1) \text{PURP}(2) [\text{USEF}(0) + \text{ROUT}(1)] + \text{DIST}(0) \text{PURP}(1) \text{USEF}(1)$$

When three logical cases are included, the instrumental use is low:

- *if there is a large distance* between the evaluator and the evaluatee, and if the evaluation is simultaneously formative and summative, *or*
- *if there is a small distance* between the evaluator and the evaluatee, and if the evaluation is summative.

In formal terms, this result may be expressed as follows:

$$\text{INST}(0) = \text{DIST}(1) \text{PURP}(2) + \text{DIST}(0) \text{PURP}(1)$$

This outcome confirms the results based on cases with positive outcomes. With respect to the influence of the institutional distance on the instrumental use of the evaluations that were examined, it may therefore be stated that a high instrumental use in the case of formative evaluations came about regardless of the distance between evaluator and evaluatee. On the other hand, summative evaluations were reliant on a large distance in order to be used instrumentally. Evaluations which pursued formative as well as summative goals benefited from a small distance between evaluator and evaluatee.

Results for "conceptual use"

As mentioned above, we shall restrict the presentation of the QCA results on conceptual use to the main findings. The analysis indicates that the combination of formative and summative objectives represented a sufficient condition for the conceptual use of the evaluations examined. The PCA, SFAO, FOT and SFOE cases are grouped together in this regard. The result can be interpreted to mean that the relatively open objectives of these evaluations played a part in favouring their conceptual use. The other two evaluations which led to results with high conceptual use (SDC and FOM) pursued formative objectives with a small institutional distance between evaluator and evaluatee. On the other hand, low conceptual use was ascertained when the purpose of the evaluation was summative or when formative goals were pursued with a large distance between evaluator and evaluatee.

Results for "process-oriented use"

With regard to the combinations of factors which favour a process-oriented effect, it can be discerned that the potential for implementation was a necessary condition for process-oriented use. If, as in the cases of PCA, SDC and SFOE, this condition was not met, no process-oriented effects came about. In addition to this factor, however, another condition also had to be met in order for high process-oriented use to occur: either the distance between evaluator and evaluatee had to be small, or the purpose of the evaluation had to be formative. Both conditions suggest that a certain proximity between evaluator and evaluatee was a necessary condition for process-oriented use.

Results for "symbolic use"

As regards the combinations of factors that support symbolic use, no conclusions can be drawn from the case studies since the sample only included one case with this type of use.

Results for the "general use" target variable

Among the ten evaluations examined, general use always came about when formative goals were pursued. If summative goals were pursued, the general use was reliant on the performance of the evaluation on a routine basis. Evaluations with formative objectives therefore offered good conditions for being used, regardless of the distance between evaluator and evaluatee. If summative goals were pursued, general use only came about (on the basis of the cases studied) if the institution had experience of dealing with evaluations.

5. Synthesis

5.1 Does distance matter?

Scriven (1997) and others have generally stressed that evaluations with a summative approach should emphasize the distance between the evaluator and the evaluatee, because acceptance of the results (and hence their utilization) are optimal in this case. On the other hand, there are interactionist approaches to evaluation (e.g. Patton 1997) advancing the view that in the case of formative evaluations, care should be taken to ensure the lowest possible distance between the evaluator and the evaluatee in order to increase the use of the evaluations. The present analysis shows that the relationship between the distance and the utilization of an evaluation appears to be more complex, and to depend on the type of utilization:

- high *instrumental use* may occur regardless of the distance. The condition for this is that formative objectives must be pursued. On the other hand, the attainment of summative objectives relies on a large distance.
- high *conceptional use* may also occur regardless of the distance. QCA identified the combination of formative and summative objectives as a sufficient condition for conceptual use.
- we only encountered a *symbolic use* of the evaluation in one of the ten case examples, from which we draw no conclusions.
- high *process-related use* only occur if the usefulness standards are respected. However, this is still not sufficient. In addition, the evaluation either had to be carried out with a small distance between evaluator and evaluatee, or it had to have a formative objective.
- Finally, *general use* always occurred when the evaluations were formative, among the cases examined. For summative evaluations, general use was reliant on the investigation being undertaken on a routine basis.

Under certain conditions, the distance between evaluators and evaluatees therefore appears to have no influence on the use of evaluations. In particular, formative objectives can be attained quite independently of the distance. On the other hand, summative objectives are generally reliant on a large distance between evaluators and evaluatees, although investigations of this sort must be carried out in a context where evaluations are part of the routine.

However, when interpreting these results, we must not neglect the fact that they are based solely on the systematic evaluation of ten case studies with QCA. Generalization is not permissible on this basis, nor is this the aim of the present article. On the contrary, the objective was to continue developing the debate (which we consider to have been neglected hitherto) about the influence of the distance between evaluators and evaluatees on the utilization of evaluations, and to formulate differentiated hypotheses which can then be tested in a subsequent qualitative stage of investigation. At the same time, some first recommendations for evaluation practice can be derived. Accordingly, the implementation of summative evaluations seems to succeed best where those responsible have a certain routine for dealing with evaluations at their disposal. This suggests an increased institutionalisation of evaluations (Sonnichsen 1999).

5.2 Does QCA provide further help?

From the viewpoint of methodology, it may be stated that Qualitative Comparative Analysis has played an important part in sharpening the focus on patterns of combination for the type of use and the influencing factors related to the environment and the evaluation process. The need for a highly systematic approach to the cases makes precise description and definition of the variables and characterizations essential. This in turn makes the case analysis transparent and contestable, meeting two fundamental conditions of scientific work (Vanderborgh et al. 2003).

Like any method used in the social sciences, QCA is nevertheless subject to severe limitations that must be fully understood before firm conclusions can be reached and recommendations formulated: firstly, it should be pointed out that QCA only refers to combinations of factors and not to connections between effects. However, by referring back to the cases after the analysis, it is possible to identify connections that help us to understand the combinations that were encountered. Most importantly of all, the results of QCA must be examined in relation to the real cases, especially if contradictory simplifying assumptions were included. QCA is primarily a valuable aid which supports systematic comparative case analysis. Nevertheless, the comparison must be restricted to a few variables, and this requires numerous simplifications. In order to analyze ten cases, for example, QCA imposes a limit of four to five influencing variables (De Meur et al. 2002). In many cases, moreover, QCA yields several solutions and it is necessary to decide which of them will be pursued further.

It is our view that QCA offers valuable support with systematic case description. The method makes precision mandatory. On the other hand, the results from QCA are difficult to interpret, and they occasionally appear to be somewhat arbitrary. For this reason, Ledermann (2004) is right to point out that QCA is most valuable if the methodology can be used as an iterative process of examination between the real cases and the results.

Andreas Balthasar, Ph.D. Political Science is founder and head of the Institute for Policy Studies Interface in Lucerne, Switzerland. He works mainly as an evaluation consultant. His activities also focus on research and consulting in the field of social and health policy. For example, he has been providing support on strategic and operational issues for Environment and Health Action Programme of the Swiss Federal Office of Public Health since 1998. He has held a lectureship at the Institute of Political Science in the University of Bern since 1987. He is President of SEVAL, the Swiss

Evaluation Society. Please send correspondence to: Institute for Policy Studies Interface, Seidenhofstrasse 12, CH-6006 Lucerne, Switzerland [email: balthasar@interface-politikstudien.ch]

References

- Alkin, M. C. (1985) *A Guide for Evaluation Decision Makers*. Beverly Hills, CA: SAGE.
- Balthasar, A. (2000) *Energie 2000. Programmwirkungen and Folgerungen aus der Evaluation*. Chur: Rüegger.
- Balthasar, A. (2003) *Die Auswirkungen des institutionellen Kontexts einer Evaluation auf die Verwendung von Evaluationsergebnissen. Erster Zwischenbericht and Fortsetzungsgesuch zuhanden des Schweizerischen Nationalfonds*. Luzern: Interface Institut für Politikstudien.
- Clarke, A. (1999) *Evaluation Research : An Introduction to Principles, Methods and Practice*. London: SAGE.
- Cronqvist, L. (2003a) 'Presentation of TOSMANA : Adding Multi-Value Variables and Visual Aids to QCA', paper presented at the COMPASS Launching Conference, 16-17 September 2003, Louvain-La-Neuve (B) and Leuven.
- Cronqvist, L. (2003b). 'Using Multi-Value Logic Synthesis in Social Science', paper prepared for the 2nd General Conference of the European Consortium for Political Research (ECPR), September 2003, Marburg (D).
- Danuser, J. and G. Regula (2002) *Evaluation der Ökomassnahmen and Tierhaltungsprogramme, Bereich Artgerechte Tierhaltung. Schlussbericht Kategorie Milchkuhe*. Bern: Bundesamt für Landwirtschaft.
- De Meur, G. and B. Rihoux (2002) *L'analyse quali-quantitative comparée. Approche, techniques et applications en sciences humaines*. Louvain-la-Neuve (B): Bruylant-Academia.
- Eidgenössische Finanzkontrolle EFK (2002) *Wirtschaftlichkeitsprüfung der Subventionen an Werkstätten für die Dauerbeschäftigung von Invaliden*. Bern: EFK.
- Eidgenössisches Personalamt EPA (2002) *Bericht an den Bundesrat über die zweite Frauenförderungsperiode in der allgemeinen Bundesverwaltung 1996 - 1999*. Bern: EPA.
- Forss, K., C. C. Rebien et al. (2002) 'Process Use of Evaluations: Types of Use that Precede Lessons Learned and Feedback', *Evaluation* 8(1): 29-45.
- Gerheuser, F. W. (1999) *Evaluation der Strategie and der Produkte des Ressorts Gewerbe von Energie 2000. Schlussbericht*. Brugg: Büro für Politikberatung and Sozialforschung.
- Greenberg, D. H. and M. B. Mandell (1995) 'Research Utilization in Policymaking : A Tale of Two Series (of Social Experiments)', in R. C. Rist (ed.) *Policy Evaluation : Linking Theory to Practice*, pp. 156-179. Aldershot: Edward Elgar
- Henry, G. T. (2000) 'Why Not Use?', in V. J. Caracelli and H. S. Preskill (eds) *The Expanding Scope of Evaluation Use. (Series: New Directions for Evaluation 88)* pp. 85-98. San Francisco, CA: Jossey-Bass.
- Henry, G. T. and M. M. Mark (2003) 'Beyond Use : Understanding Evaluation's Influence on Attitudes and Actions', *American Journal of Evaluation* 24(3): 293-314.
- Huberman, M. and M. Gather Thurler (1991) *De la recherche à la pratique. Eléments de base*. Berne: Peter Lang.
- Jaggi, K. and A. Näf (2000) *Evaluation der Rückkehrberatung im Asylbereich. Schlussbericht zuhanden des Bundesamts für Flüchtlinge (BFF) and der Internationalen Organisation für Migration (IOM)*. Bern: BFF; IOM.
- Joint Committee on Standards for Educational Evaluation (2000) *Handbuch der Evaluationsstandards : die Standards des "Joint Committee on Standards for Educational Evaluation"*. Opladen: Leske + Budrich.
- Kingdon, J. W. (1984) *Agendas, Alternatives, and Public Policies*. Boston: Little, Brown and Company.
- Ledermann, S. (2004) *Wozu denn wissenschaftlich? Untersuchung zur Verwendung von Evaluationen in der schweizerischen Entwicklungszusammenarbeit*. Bern: Institut für Politikwissenschaft der Universität Bern.
- Lehmann, L. and A. Balthasar (2004) *Quality Assessment of SDC's External Evaluation Reports, Commissioned by the Evaluation & Controlling Division of the Swiss Agency for Development and Cooperation (SDC)*. Lucerne: Interface Institute for Policy Studies.
- Lester, J. P. and L. J. Wilds (1990) 'The Utilization of Public Policy Analysis: A Conceptual Framework', *Evaluation and Program Planning* 13: 313-319.
- Mark, M. M. and G. T. Henry (2004) 'The Mechanisms and Outcomes of Evaluation Influence', *Evaluation* 10(1): 35-57.

- Owen, J. M. (1993) *Program Evaluation : Forms and Approaches*. St Leonards, NSW: Allen and Unwin.
- Parlamentarische Verwaltungskontrollstelle PVK (2000) *Das Kartellrecht: Standortbestimmung. Schlussbericht zuhanden der Geschäftsprüfungskommission des Nationalrates*. Bern: PVK
- Patton, M. Q. (1997) *Utilization-Focused Evaluation: The New Century Text*. Thousand Oaks: SAGE.
- Patton, M. Q. (1998) 'Discovering Process Use', *Evaluation* 4(2): 225-233.
- Ragin, C. C. (1987) *The Comparative Method. Moving beyond qualitative and quantitative strategies*. Berkley, CA: University of California Press.
- Rogers, P. J. and G. Hough (1995) 'Improving the effectiveness of evaluations: Making the link to organizational theory', *Evaluation and Program Planning* 18: 321-32.
- Sardi, M. and A. Ensmann (2002) *L'Evaluation de la Campagne nationale de prévention de la grippe 2001*. Carouge, GE: Institut érasme.
- Schweizerischer Bundesrat (2002) *Bericht über die Verkehrsverlagerung von 27. März 2002 (Verlagerungsbericht 2002). Bericht des Bundesrates an die parlamentarischen Kommissionen*. Bern: Departement für Umwelt, Verkehr, Energie und Kommunikation
- Scriven, M. (1997) 'Truth and Objectivity in Evaluation', in E. Chelimsky and W. R. Shadish (eds) *Evaluation for the 21st Century – A Handbook*, pp. 477-500. Thousand Oaks, CA: SAGE
- Shulha, L. M. and J. B. Cousins (1997) 'Evaluation Use : Theory, Research, and Practice since 1986', *Evaluation Practice* 18(3): 195-208.
- Sonnichsen, R. C. (1999) 'Building Evaluation Capacity Within Organisations', in R. Boile and D. Lemaire (eds) *Building effective evaluation capacity: lessons from practice*, New Jersey: Transaction Publisher: 53-73.
- Vanderborgh, Y. and S. Yamasaki (2003) *The Problem of Contradictory Simplifying Assumptions in Qualitative Comparative Analysis (QCA)*. Magdeburg (Germany): European Consortium for Political Research.
- Varone, F., B. Rihoux et al. (2005) 'A New Method for Policy Evaluation? Longstanding challenges of policy evaluation and the possibilities of Qualitative Comparative Analysis (QCA)', in B. Rihoux and H. Grimm (eds) *Innovative comparative methods for policy analysis and evaluation*, New York: Springer/Kluwer.
- Vedung, E. (1997) *Public Policy and Program Evaluation*. New Brunswick, NJ: Transaction.
- Vedung, E. (1999) *Evaluation im öffentlichen Sektor*. Wien: Böhlau Verlag.
- Weiss, C. H. (1977) 'Research for Policy's Sake: The Enlightenment Function of Social Research', *Policy Analysis* 3: 531-545.
- Weiss, C. H., E. Murphy-Graham and S. Brikeland (2005) 'An Alternate Route to Policy Influence : How Evaluations Affect D. A. R. E.', *American Journal of Evaluation* 26(1): 12-30.
- Wholey, J. S. (1994) 'Assessing the Feasibility and Likely Usefulness of Evaluation', in J. S. Wholey, H. P. Hatry and K. E. Newcomer (eds) *Handbook of Practical Program Evaluation*, pp. 15-39. San Francisco, CA: Jossey-Bass.
- Widmer, T., Ch. Landert and N. Bachmann (2000) 'Evaluation Standards Recommended by the Swiss Evaluation Society (SEVAL)', in C. Russen (ed.) *The Program Evaluation Standards in International Settings*, pp. 81-102. Kalamazoo, MI: Western Michigan University.
- Widmer, T., E. Rüegg et al. (2001) *Stand and Aussichten der Evaluation beim Bund. Schlussbericht zuhanden der Schweizerischen Bundeskanzlei vom 9. März 2001*. Zürich: Institut für Politikwissenschaft der Universität Zürich.
- Widmer, T. and P. Neuenschwander (2004) 'Embedding Evaluation in the Swiss Federal Administration: Purpose, Institutional Design and Utilization', *Evaluation* 10(4): 388-409.
- Williams, K., B. de Laat et al. (2002). *The Use of Evaluation in the Commission Services. Final Report*. Paris: Technopolis France.

ⁱ This is a revised version of a paper presented at the European Evaluation Society 6th Biennial Conference in Berlin, Germany, 30 September – October 2, 2004. The author would like to thank Manuela Oetterli for her support as well as the three anonymous reviewers for their helpful comments.

ii The analysis was carried out using the TOSMANA software, Version 1.102 (Cronqvist 2003; Cronqvist 2003b). The TOSMANA software can be downloaded free of charge from the following website: www.tosmana.org.