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

ANDREAS BALTHASAR
Interface Institute for Policy Studies, Switzerland

This article presents some of the results from a study in progress, focusing on the influence of the institutional distance between evaluators and evaluees 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 carried out in different institutional contexts, with widely varying institutional distances between evaluators and evaluees. '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 institutional distance between evaluators and evaluees has no influence on the use of evaluations. In particular, formative objectives can be achieved quite independently of 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 possible 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 institutional distance between evaluators and evaluees on the utilization of evaluations.

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

1. Introduction and Question to be Answered

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 (Weiss, et al. 2005; Widmer et al., 2004) and the mechanisms governing the outcomes of evaluation (Henry et al., 2003), only a few contributions deal with the influence of evaluation-specific institutional design on utilization (Rogers et al., 1995; Williams et al., 2002). However, this

design normally determines the institutional distance between evaluators and evaluees. This distance is a highly important factor influencing the utilization of evaluations. In an article in this journal, Widmer and Neuenschwander (2004) presented some insights into the relationship between utilization and institutional design in analysing Federal Swiss evaluation practice. They concluded that the lack of use of an evaluation is often due to a bad match between the institutional setting and the purpose of an evaluation. The present article develops this issue further, focusing on the influence exerted by the institutional distance between the evaluator and the evaluee 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-based comparisons.

The first section of this article outlines our analytical framework, followed by descriptions of the method and database. After presenting the results of the analysis, the article concludes 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 use – instrumental, conceptual, symbolic and process-related – 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* used here. This classification is very similar to a distinction proposed by Vedung (1997, 1999) which was recently used by Widmer and Neuenschwander (2004), and also by 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 that 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 to confirm their perspective and to obtain legitimation (Weiss et al., 2005: 13). Referring to Owen (1993), 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 (1997, 1998) and 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 equivalent 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 that influence the use of evaluations*. Alkin (1985), for example, differentiates 50 factors that help determine the use of evaluation. Huberman and Gather Thurler (1991) identify 47 factors of influence that have proven significant in terms of evaluation implementation. Shulha and Cousins (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 and Mark, 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 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 evaluee, 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 when and if suitable opportunities arise to do so (Kingdon, 1984). Lester and Wilds (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 (including, for example, work by Greenberg and Mandell, 1995) also suggest that factors relating to the way an evaluation is undertaken are very important for the use of the evaluation results. Lester and Wilds (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 (2000). These standards are based on a fundamental assumption: in order to be considered of high quality, the process of an evaluation must simultaneously satisfy requirements regarding usefulness, implementability, correctness and accuracy (cf. Widmer et al., 2000).

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

3. Method

3.1. Database

A three-stage process was used to select the case studies. The first stage comprised assembling 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).

The second stage selected 10 of the 32 Federal authorities who had carried out at least one evaluation. Two criteria were taken into account for selection purposes. First, the Federal authorities had to fall into different institutional contexts, in order to ensure maximum variability of the influencing variables. 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). Second, the aim was for the Confederation's control bodies and all seven Departments of the Confederation to be represented in the case studies. Table 1 briefly presents the 10 institutions chosen for further investigation, indicating the institutional distance between evaluators and evaluees in each context. This indication is based on expert interviews and the author's assessment of the institutional context key influencing factors described.

The third stage comprised a random process to identify one evaluation from each of the selected authorities, to be examined in an in-depth case study. The random selection process was used here as this was considered to be the best way of finding a typical case for each of the 10 selected institutional settings. Table 2 briefly presents the evaluation 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).

Table 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)	It 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 30 years. Nowadays, evaluation functions are embedded in Program Cycle Management (Lehmann and Balthasar, 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 programmes under examination. Financing is provided from the budget for the programme.	Small
Swiss Federal Office of Public Health (SFOPH)	Evaluation began to play a part at the SFOPH in the late 1980s (Widmer and Neuenschwander, 2004). Today the directorate-level Centre de Competences 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)	This is the department's internal audit body and 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 and has no special budget for this activity.	Quite small

(Table 1 continued)

Institution	Short description of the institution	Distance
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 e.g. in connection with combating epidemics. Ex-post policy evaluations as such are rare. The SFVO does not have a central unit with responsibility for evalua-tions. 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: 201–20). 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

Table 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	Rather low

(Table 2 continued)

(Table 2 continued)

Institution	Short description of the evaluation	Assessment of utility
	that responsibility for the permanent workshops had to be shifted from the Confederation to the Cantons as part of a redistribution of tasks.	
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 Competences on 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 and Ensmann, 2002). Most of the suggested	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 and Näf, 2000). The evaluators were designated without obtaining different offers. A group of responsible FOM staff provided intensive follow-through for 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.	Quite high
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

(Table 2 continued)

(Table 2 continued)

Institution	Short description of the evaluation	Assessment of utility
Swiss Federal Veterinary Office (SFVO)	A comparatively new methodological approach was applied in the evaluation of a measure to promote the keeping of dairy-cows in a manner suitable to the species (Danuser and Regula, 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 many outcomes from the analysed 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 10 cases, it was necessary to identify factors and categories on the basis of the analytical framework described.

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 factors on which the type and intensity of use depend. For each of the five factors, two levels of intensity of use 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.

Influencing factors arising from the institutional context   In order to identify the relevant factors that can define the distance between evaluators and evaluees, we analysed the available literature (Rogers and Hough, 1995; Widmer and Neuenschwander, 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   It was assumed that the institutional distance between evaluators and evaluees is greater if a unit located outside of line responsibility for the programme is in charge of administering the evaluation than if the programme director handles the evaluation themselves.

The existence of an evaluation budget   It was further assumed that the institutional distance between evaluators and evaluees is greater if the evaluations are not financed from the funds for the evaluated action.

The way in which the evaluation is triggered   It was assumed that in a case where an evaluation unit or a decision-maker can initiate the evaluation, the institutional distance between evaluators and evaluees is greater than in a case where only the individuals responsible for the programs or actions can do so.

Allocation practice   It was assumed that the institutional distance between evaluators and evaluees is greater if there is a competitive procedure (invitation to tender) for assigning evaluations rather than where projects are directly awarded to evaluators who appear suitable.

Publication practice   We assumed that the institutional distance between evaluators and evaluees is greater if the reports are consistently published. 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   It is assumed that the institutional distance between evaluators and evaluees is greater for external evaluations. If external evaluation is controlled by a strong internal steering group, it does not necessarily lead to a large distance between evaluators and evaluees. However, we assume that, in any case, distance is greater when an external evaluator is present compared with self-evaluation.

The key influencing factor of 'institutional distance between the evaluator and the evaluee' (DIST) was formed, on the basis of these six factors. This value expresses whether – and how – the evaluees are able to influence the evaluation. We assume that the evaluee's influence on the evaluation increases as the

distance between the evaluator and the evaluee 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 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 routinely undertaken. Two levels are allowed (yes/more yes, more no/no).

Influencing factors arising from the evaluation process   The analysis uses the 'usefulness' (USEF) variable to represent the possible 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 whether there 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 et al., 2000). The USEF variable includes two levels (yes/more yes, more no/no).

The data were collected in two stages. During the first stage, the listed variables for each of the 10 selected evaluations were recorded during a semi-structured 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 during the second stage, interviews were conducted with potential users of the evaluation. 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 (De Meur and Rihoux, 2002; Ragin, 1987; Varone et al., 2005). QCA is based on Boolean algebra and is used to search systematically for commonalities and differences between cases. QCA emphasizes the importance of combinations of causes to bring about specific results. The method is based upon the assumption that observable effects depend on a combination of factors. Consequently, it does not involve individual explanatory factors but rather – according to the logic of QCA – the combinations in which these factors occur, as a way of explaining particular results.[1] QCA analysis was carried out in three stages (De Meur and Rihoux, 2002; Ledermann, 2004).

During the first stage, the configurations of factors for the 10 cases were 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. During the second stage, 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. During the third stage of the analysis, the factor combination patterns were simplified even further. In this step, those configurations that were theoretically possible but which were not observed in the 10 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. 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 actual cases (Vanderborght and Yamasaki, 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. It begins by considering the configurations of variables for the 10 cases that were studied.

Table 3. Influencing factors and target factors

	Influencing Factors				Target Factors				
	DIST	PURP	ROUT	USEF	INST	CONC	PROC	SYMB	GENE
PCA/SFOE	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

Table 3 shows that 10 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), there are six combinations with high-intensity use. Five cases show highly intensive process-oriented use of the evaluation (PROC). As regards the fourth target variable – the symbolic use of evaluation (SYMB) – there is one configuration that produced high-intensity use. At a general level, six evaluations were regarded as useful (GENE).

For the second and third stages of the analysis, the focus will initially be limited to those cases that were observed in reality. Subsequently the 'logical cases' are incorporated into the analysis. 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, the analysis for 'instrumental use' is presented in detail. The results for the other elements of the analysis will merely be summarized.

Results for 'Instrumental Use

Analysis of the five actual cases with a positive outcome shows that instrumental use of the results from the evaluations is *high* if:

- there was a *large distance* (DIST 1) between the evaluator and the evaluee, the evaluation was carried out on a routine basis (ROUT 1), there was potential for implementation (USEF 1) and the evaluation was either formative (PURP 0) or summative (PURP 1), *or*
- there was a *small distance* (DIST 0) between the evaluator and the evaluee, *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:

INST(1) = DIST(1) ROUT(1) USEF (1) [PURP(0) + PURP(1)] + DIST(0) [PURP(0) ROUT(1) USEF(0) + PURP(0,2) ROUT(0) USEF(1)]

- When 11 logical cases are included, the instrumental use is high: if the evaluation was formative (PURP 0) *or* if the evaluation was carried out with 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:

INST(1) = PURP(0) + DIST(0) PURP(2) + DIST(1) PURP(1)

Thus, 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 evaluee 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 evaluee, the instrumental use of evaluations with summative objectives was reliant on a large distance between evaluator and evaluee.

In the five actual cases with low instrumental use, the instrumental use is *low* if*:

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

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

INST(0) = DIST(1) PURP(2) [USEF(0) + ROUT(1)] + DIST(0) PURP(1) USEF(1)

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

- there was a *large distance* between the evaluator and the evaluee, and the evaluation was simultaneously formative and summative, *or*
- there was a *small distance* between the evaluator and the evaluee, and the evaluation was summative.

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

INST(0) = DIST(1) PURP(2) + DIST(0) PURP(1)

This outcome confirms the results of the analysis of cases with positive outcomes. With respect to the influence of institutional distance on the instrumental use of the evaluations that were examined, a high instrumental use in the case of formative evaluations came about regardless of the distance between evaluator and evaluee. On the other hand, where the evaluation was summative, a large distance was required in order for the evaluation to be used instrumentally. Evaluations that pursued formative as well as summative goals benefited from a small distance between evaluator and evaluee.

Results for 'Conceptual Use'

As has been mentioned, the presentation of the QCA results on conceptual use will be restricted to the main findings. The analysis indicates that the combination of formative and summative objectives was a sufficient condition for the conceptual use of the evaluations examined. The PCA, SFAO, FOT and SFOE cases are grouped together for the purpose of this analysis. 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 evaluee. On the other hand, low conceptual use was achieved when the purpose of the evaluation was summative or when formative goals were pursued with a large distance between evaluator and evaluee.

Results for 'Process-Oriented Use'

With regard to the combinations of factors that favour process-oriented use, the analysis suggests that the potential for implementation was a necessary condition for this type of use. If, as in the cases of PCA, SDC and SFOE, this condition was not met, no process-oriented effects occurred. 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 evaluee had to be small, or the purpose of the evaluation had to be formative. Both conditions suggest that a degree of proximity between evaluator and evaluee was a necessary condition for process-oriented use.

Results for 'Symbolic Use'

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

Results for 'General Use'

Among the 10 evaluations examined, general use always occurred when formative goals were pursued. If summative goals were pursued, general use was reliant on

whether the evaluation was routine or not. Evaluations with formative objectives therefore offered good conditions for being used, regardless of the institutional distance between evaluator and evaluee. If summative goals were pursued, general use only occurred (on the basis of the cases studied) if the institution had experience of dealing with evaluations.

5. Synthesis

5.1. Does Institutional Distance Matter?

Scriven (1997) and others have stressed that evaluations with a summative approach should emphasize the institutional distance between the evaluator and the evaluee, because acceptance of the results (and hence their utilization) is optimal in this case. On the other hand, there are interactionist approaches to evaluation (e.g. Patton, 1997) advancing the view that, when an evaluation is formative, care should be taken to ensure the least possible distance between the evaluator and the evaluee 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 distance. The pre-condition for this is that formative objectives are being 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 10 cases, from which we could draw no conclusions.
- High *process-related use* may only occur if the potential for implementation is attributed to the evaluation. However, this is still not sufficient. In addition, the evaluation either had to be carried out with a small distance between evaluator and evaluee, 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 institutional distance between evaluators and evaluees therefore appears to have no influence on the use of evaluations. In particular, formative evaluation objectives can be attained quite independently of distance. On the other hand, summative evaluation objectives are generally reliant on a large distance between evaluators and evaluees, although investigations of this sort must be carried out in a context where evaluation is routine.

However, when interpreting these results, it must be borne in mind that they are based solely on the systematic evaluation of 10 case studies with QCA. Generalization is not permissible on this basis, nor was this the aim of this article.

On the contrary, the aim was to continue developing the debate (which has been neglected hitherto) about the influence of the institutional distance between evaluators and evaluees on the utilization of evaluations, and to formulate differentiated hypotheses that can then be tested at a subsequent qualitative stage of investigation. However, some preliminary recommendations for evaluation practice can be derived. Thus, summative evaluations seem to succeed best where those responsible have a certain routine for dealing with evaluations at their disposal. This would follow from increased institutionalization of evaluations (Sonnichsen, 1999).

5.2. Does QCA Provide Further Help?

Regarding methodology, it appears 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 analysis (Vanderborght and Yamasaki, 2003).

Like any method used in the social sciences, QCA is subject to severe limitations that must be fully understood before firm conclusions can be reached and recommendations formulated. First, 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 actual cases, especially if contradictory simplifying assumptions were included. QCA is primarily a valuable aid to support systematic comparative case analysis. Nevertheless, the comparison must be restricted to a few variables, and this requires numerous simplifications. In order to analyse 10 cases, for example, QCA imposes a limit of four to five influencing variables (De Meur and Rihoux, 2002). In many cases, moreover, QCA yields several solutions and it is necessary to decide which of them will be pursued further.

To conclude, 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 part of an iterative process moving back and forth between actual cases and the results of analysis.

Notes

This is a revised version of a paper presented at the European Evaluation Society 6th Biennial Conference in Berlin, Germany, 30 September–2 October 2004. The author

Evaluation 12(3)

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

References

Alkin, M.C. (1985) *A Guide for Evaluation Decision Makers*. Beverly Hills, CA: SAGE.

Balthasar, A. (2000) *Energie 2000: Programmwirkungen und Folgerungen aus der Evaluation*. Chur: Rüegger.

Balthasar, A. (2003) *Die Auswirkungen des institutionellen Kontexts einer Evaluation auf die Verwendung von Evaluationsergebnissen: Erster Zwischenbericht und Fortsetzungsgesuch zuhanden des Schweizerischen Nationalfonds.* Lucerne: 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 Sept., 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), Sept., Marburg (D).

Danuser, J. and G. Regula (2002) *Evaluation der Ökomassnahmen und Tierhaltungsprogramme, Bereich Artgerechte Tierhaltung: Schlussbericht Kategorie Milchkühe*. Berne: 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.* Berne: EFK.

Eidgenössisches Personalamt EPA (2002) *Bericht an den Bundesrat über die zweite Frauenförderungsperiode in der allgemeinen Bundesverwaltung 1996–1999*. Berne: 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 und der Produkte des Ressorts Gewerbe von Energie 2000: Schlussbericht*. Bruges: Büro für Politikberatung und 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. 156–79. 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*. New Directions for Evaluation 88: 85–98. San Francisco: 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) und der Internationalen Organisation für Migration (IOM)*. Berne: BFF and 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 & Co.

Ledermann, S. (2004) *Wozu denn wissenschaftlich? Untersuchung zur Verwendung von Evaluationen in der schweizerischen Entwicklungszusammenarbeit*. Berne: Institut für Politikwissenschaft der Universität Bern.

Lehmann, L. and A. Balthasar (2004) *Quality Assessment of SDC's External Evaluation Reports, Commissioned by the Evaluation and 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: 313–19.

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 & Unwin.

Parlamentarische Verwaltungskontrollstelle PVK (2000) *Das Kartellrecht: Standortbestimmung: Schlussbericht zuhanden der Geschäftsprüfungskommission des Nationalrates*. Berne: PVK.

Patton, M.Q. (1997) *Utilization-Focused Evaluation: The New Century Text*. Thousand Oaks, CA: SAGE.

Patton, M.Q. (1998) 'Discovering Process Use', *Evaluation* 4(2): 225–33.

Ragin, C.C. (1987) *The Comparative Method: Moving beyond Qualitative and Quantitative Strategies*. Berkeley, 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 érasm.

Schweizerischer Bundesrat (2002) *Bericht über die Verkehrsverlagerung von 27. März 2002 (Verlagerungsbericht 2002): Bericht des Bundesrates an die parlamentarischen Kommissionen*. Berne: 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*, pp. 53–73. New Brunswick, NJ: Transaction.

Vanderborght, 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 and A. Marx (2006) '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*, pp. 213–36. New York: Springer/Kluwer.

Vedung, E. (1997) *Public Policy and Program Evaluation*. New Brunswick, NJ: Transaction.

Vedung, E. (1999) *Evaluation im öffentlichen Sektor*. Vienna: Böhlau Verlag.

Weiss, C.H. (1977) 'Research for Policy's Sake: The Enlightenment Function of Social Research', *Policy Analysis* 3: 531–45.

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: Jossey-Bass.

Widmer, T. and P. Neuenschwander (2004) 'Embedding Evaluation in the Swiss Federal Administration: Purpose, Institutional Design and Utilization', *Evaluation* 10(4): 388–409.

Widmer, T., C. 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 and P. Neuenschwander (2001) *Stand und Aussichten der Evaluation beim Bund: Schlussbericht zuhanden der Schweizerischen Bundeskanzlei vom 9. März 2001*. Zürich: Institut für Politikwissenschaft der Universität Zürich.

Williams, K., B. de Laat and E. Stern (2002) *The Use of Evaluation in the Commission Services: Final Report*. Paris: Technopolis France.

ANDREAS BALTHASAR is the 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. 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]