A Theoretical Framework and Grid for Analysis of Programme-evaluation Practices

MICHEL DEMARTEAU

Public Health Institute of Province de Hainaut, Belgium

This article proposes a theoretical framework and a grid to unify models and to describe programme evaluation. It identifies three components in the evaluation process: the information-collecting process; the value-judgement product; and the use of the latter in decision making. It defines three differentiation axes (or dimensions): the social enunciation, the external or internal position of actors, and the formal or informal character of the evaluation components. The relevance for the practitioner and the researcher is discussed and briefly illustrated with two examples from the health promotion sector: a project selection process and a retrospective study of programme-evaluation practices.

KEYWORDS: evaluation practice; health promotion; programme evaluation; public health; theoretical model

Introduction

The origin of this work stems from the need to clarify the work of evaluation counsellors. This has been developed by the French Community of Belgium, a political body within the federal structure of Belgium, in charge of education, culture or health promotion and education for the French-speaking population. The aim is to contribute to the quality of health-education programmes (Demarteau, 1994). In this approach, the evaluation counsellor's role is not to evaluate a programme, but to support programme managers to conceive, formulate and exploit their own programme evaluations. He/she has therefore to depend on already existing evaluation practices. So the quality of counselling is dependent on the quality of description and understanding of these practices. Contributing to the quality of programmes requires: identifying the current practices and proposing intervention models in evaluation that would be better adapted to the needs of the users in the field.

How can diversity in practices be described? On what basis can practices be compared? Are there best practices or evidence-based practices in evaluation? Researchers and programme managers consulting the vast literature devoted to

programme evaluation face a diversity and complexity of vocabulary and theoretical models. Evaluation is a large, but not unified theoretical area. No single theory specific to evaluation is available to describe, explain and predict all types of evaluation activities (Alkin and Ellett, 1990). Nevertheless, a more comprehensive evaluation theory would be useful for all those involved in evaluation, particularly for programme managers, advisors and counsellors.

After briefly reviewing previous attempts to unify theoretical models and to describe programme-evaluation practices, this article proposes a theoretical framework and grid for this purpose. Finally, the relevance for the practitioner and the researcher is discussed and briefly illustrated with two examples from the health promotion sector: a project selection process and a retrospective study of programme-evaluation practices.

Overview of the Situation

Despite an extensive review of the literature, in fields such as education, social work and public health, we were unable to identify a general theory of programme evaluation that would take all the models constructed by evaluators into account and integrate external and internal evaluation practices.

But we have found three distinct attempts to analyse evaluation models (theorized practices of evaluation professionals) and practices. The first is taxonomical, mainly descriptive. The second tries to analyse theoretical models by using scientific methods. The third addresses particular areas of research on evaluation practice.

Taxonomies

While there is no general theory of evaluation, there are many taxonomies of evaluations. Authors have attempted to categorize and to synthesize the different theories and practices, to order them and to allow comparisons: Alkin and Ellett (1990); Guba and Lincoln (1989); House (1978); Monnier (1987); Stufflebeam and Webster (1980); Worthen (1990); and many others. These classifications generally reflect two main designs (Williams, 1989): an analysis of the apparent structure and an analysis of the underlying structure.

The analysis of apparent structure puts together models and practices based on similarities. They are differentiated on distinct and juxtaposed characteristics and not on the variation of a common characteristic.

The other design compares the different models and practices on underlying dimensions and principles. Dimensions are generated *a priori*, then they are classified according to the former. For example, House (1978) analyses evaluation models on underlying philosophical concepts.

This taxonomical current is essentially qualitative, based on the judgements and perceptions of the authors.

Research on Evaluation Models

The second attempt at analysing evaluation models and practices follows a particular methodology.

Thus, Alkin and Ellett (1979) studied the theory development process with a naturalistic, qualitative methodology. They focus on the historical and contextual dimensions of evaluation practice and of theory.

In another attempt at theoretical model analysis, Williams (1989) studied differences between theoretical models and practices based on these models, using a quantitative methodology. Fourteen evaluation leaders answered a questionnaire on the similarity of their theory with those of other evaluators and on the characteristics of their current practices. Answers were processed using multi-dimensional scales to provide maps of underlying dimensions of theories and practices. These relied on the subjective judgement of authors. Results indicate that perceived differences between theories are more important and more complex than perceived differences between practices themselves. The mapping technique makes evident a continuum differentiating evaluators, especially on the 'method' and the 'utilization' dimensions. This study also underlines the likely difference between theories and their related practices.

The aim of this study was to strengthen the image of a plural approach to evaluation through the diversity of theoretical positions but it tells us nothing about their actual practices.

Research on Evaluation Practices

Beyond the analysis of models, some researchers have shown an interest in analysing evaluation practices. From the late 1970s onwards, empirical studies on the use of evaluation results began (Cousins and Leithwood, 1986). Studies of explanatory factors in the use of evaluation have mostly dealt with the relationships between external evaluations, undertaken by professionals, and political and administrative decisions. However, whilst these studies are related to utilization practices, there has been little research on the practices themselves.

Marsch et al. (1981) have used the 'standards' proposed by the 'Joint Committee on Standards for Educational Evaluation' for the study of the education evaluation practices in California. When using this approach, practices are not judged with respect to their actual effectiveness, but are compared to a methodological ideal. Invoked criteria are then of an extrinsic nature (i.e. time, policy and competence). The nature of evaluations undertaken and the reasons for their appearance are not explained.

In another interesting study, Shadish and Epstein (1987) pointed to the abundance of theoretical – often contradictory – advice in the area of evaluation. In their view, the studies available remain centred on the description of methods used without attempting to identify the factors that determine how the evaluators conceive and achieve their work. They raised the question of evaluator practices. These authors defined evaluation practices on the basis of a large survey of members of two professional evaluators' associations. They show the great variety of professional practices. Two main factors are suggested to explain this diversity: evaluation training and the professional situation of evaluators. The study remained descriptive of the practices of professional evaluators.

Theoretical Basis of the Analysis of Programme-evaluation Practices Framework

As this overview of the literature shows, there is little research on evaluation and even less devoted to the study of evaluation practices. The aim of this study therefore was to construct a theoretical framework that would allow the identification, description and improvement of current understanding of evaluation practices. The first step is to form a definition of evaluation.

A Definition of Programme Evaluation with Three Components

Barbier (1990) defined evaluation by the presence of a value judgement. But, this alone was insufficient for him. He distinguished three forms of value judgement: implicit, spontaneous and instituted. The difference between these three forms is based on the implicit or explicit nature of three components of evaluation: the process of evaluation, the value judgement and the use made of this value judgement.

‘Implicit’ Evaluation In this form, only the use of the value judgement is explicit. ‘The value judgement is made visible only through its effects’¹ (Barbier, 1990: 32). This first manifestation of evaluation can be seen as a universal social act. Every pursuit of an action implies an evaluation of its value.

‘Spontaneous’ Evaluation ‘The value judgement is only made visible through its statement, its formulation’ (Barbier, 1990: 33). The value judgement and its use are explicit, but it is neither socially organized, nor expressed by persons socially recognized to do so.

‘Instituted’ Evaluation ‘The value judgement is made entirely visible as a product of a specific social process whose main stages are likely to be observed’ (Barbier, 1990: 34). For Barbier, ‘instituted’ evaluation involves three components: use of well-defined instruments and methodology; socially authorized persons to conduct it; and formalized results to be used.

The theoretical framework that we propose takes its inspiration from the concepts developed by Barbier.

We will define programme evaluation as an information-collection process leading to a value judgement, which will be used in decision-making, leading to action.

Evaluation involves three components: the process, the product and its use. All evaluation is defined by the presence of these three components. In order to understand programme-evaluation practices, it is necessary to consider evaluation from these three perspectives: the process of information collection; the value judgement that is the product of this process; and finally the use of the value judgement in the decision making leading to action. These three components also cover three evaluation issues: to know, judge and decide.

This evaluation definition based on three components is strengthened by the theoretical works of Nevo (1986) and Alkin and Ellett (1990).

Nevo (1986) reviewed definitions of education evaluation and identified three main trends. The first and most frequent trend defines evaluation as the measure of the value of an object, such as a programme. The second, less frequent, trend emphasizes the collection of information. Evaluation is a systematic investigation into the progress and consequences of a programme. The third tendency defines evaluation as a means to provide the information to make decisions.

Alkin and Ellett (1990) proposed the following three dimensions to characterize and differentiate between different models of evaluation: methodology, the definition of, and the functions attributed to evaluation. Methodology allows us to describe and explain the phenomena studied. The definition given to evaluation clarifies the importance of the value judgement and the way to attribute a value to an object. Functions enable us to specify the use of evaluation. The originality of these contributions lies in the fact they have shown that every evaluation model consists of these three dimensions even though they place different emphases on one or other of them.

Whilst we have a broad definition of programme evaluation, such a definition is still insufficient to accurately describe all types of evaluation practice.

Three Axes of Differentiation

Three important characteristics of evaluation practice can be identified: a social dimension, a strategic dimension and a technical one. These three dimensions provide a way to differentiate between practices while preserving the idea of a unity of structure common to all evaluations.

The Social Dimension or the Implicit–Explicit Axis Evaluation is accepted by most evaluators as a social practice. This social practice is expressed by the social position and the interests of the different actors in an evaluation. More generally, the social dimension is expressed first by the enunciation – a form of socialization – of one or the other of the evaluation components.

The theoretical framework proposed here is based upon the concept of explicitation (implicit–explicit differentiation) developed by Barbier (1990). But this concept has been given a more general scope. The enunciation of one or more of these components signifies the social function that is attributed to evaluation and determines its practice. This enunciation is the decisive element that will allow differentiation between multiple evaluation practices.

The implicit–explicit axis applied to the three components of evaluation provide a set of eight possible types of programme evaluation that could serve to describe evaluation practice (see Figure 1).

The Strategic Dimension or the Internal–External Axis Evaluation plays a role in the power games between actors engaged in organized activities along the lines of the model proposed by Crozier and Friedberg (1977). Shadish and Epstein (1987) also show that the professional situation of the evaluators can clearly make a difference to their professional practice.

We have attempted to integrate this strategic dimension by identifying the internal or external position of the actors of the different components of

T0: The 'Implicit' Evaluation

T1: The 'Decision' Evaluation (explicit decision)

T2: The 'Expert' Evaluation (explicit judgement)

T3: The 'Scientific' Evaluation (explicit process)

T4: The 'Control' Evaluation (explicit judgement and decision)

T5: The 'Conclusion' Evaluation (explicit process and judgement)

T6: The 'Process-Decision' Evaluation (explicit process and decision)

T7: The 'Explicit' Evaluation (explicit process, judgement and decision)

Figure 1. The Eight Types of Evaluation

evaluation, as compared to those in charge of the programme implementation and management.

The process of evaluation, the value judgement and the utilization of evaluation can each be considered as internal or external depending on the status of the person(s) in charge of the evaluation. It is internal when such persons belong to the management team, and external otherwise. Obviously any evaluation component cannot have an internal or external position without being explicated (social dimension axis).

The evaluation is explicit and internal if the programme manager collects information, judges the value and takes a decision by an explicit and recognized process. Each component can also be carried through by an external actor. The position of the different actors, their part in the programme, their 'specialization' allow us to understand the apparent discontinuity of an evaluation activity. This separation of functions is often all the greater when programmes are large scale and mobilize many people.

The Technical Dimension or the Informal-Formal Axis The third axis of differentiation is the technical (or methodological) dimension of evaluation. This remains central in speeches and debates about evaluation and therefore needs to be taken into account in any comprehensive model of evaluation practice. The methodological dimension applies most often to the process of evaluation. The concept of 'scientific' evaluation is highlighted through this dimension, often identified with an experimental model.

In order to generalize this methodological dimension, we define it by the degree of formalization of the evaluation components. Each component must have been explicated beforehand.

In our theoretical grid, an evaluation process is formal if it involves indicators, a method and tools of information collection. The value judgement is considered as formalized when the criteria, norms, values on which the judgement bears are defined. The formalization of the utilization of evaluation corresponds to the direct use of the value judgement in a formal system, for example within the framework of a board of administrators.

These social, strategic and technical dimensions are distinct. The first step is the explication and social recognition of a component. It becomes formal if a methodology is followed. But strategic and formal dimensions are not tied: external experts may work without formal methodology.

The Grid for Analysis of Programme-evaluation Practices

The combination of the three evaluation components (process, judgement, and decision) and its three dimensions (social, strategic and technical) provides an extensive theoretical grid that can differentiate between 125 different modes of programme-evaluation practice. The first is the totally implicit type of evaluation (Type 0, Figure 1) where each component is implicit and therefore, in principle, hardly perceptible. Each explicit component from the seven other types of evaluation can take four positions: formal or informal and internal or external (see Figure 2). The 125th mode is the totally explicit type (Type 7, Figure 1) where each component is external and formal.

We argue that the grid for Analysis of Programme-Evaluation Practices (APEP) allows the description of practices in the field by programme managers or evaluation professionals. Despite the diversity of the terminology used, there remains one single activity of programme evaluation whose components are made visible or not depending on the way social, strategic and technical dimensions are configured.

1. Are the components explicit or implicit?
2. Are the explicit components external or internal?
3. Are the explicit components formal or informal?

Example : 'formal external explicit Decision': one mode for 'Decision' evaluation (type 1)

Components: P = information Process; J = value Judgment; D = Decision making

Figure 2. Three Steps to Identify an Evaluation Mode

In our view, the differentiation of programme evaluation practices happens around three types of socialization: enunciation, externalization and formalization. The complexity and the duration of a programme may lead to a division of labour and to a specialization that implies some externalized practice. This specialization goes together with a more frequent enunciation necessary to justify or account for the completed work, either to peers, partners, fundraisers or the leaders of the organization or to the community. Formalization is also associated with this process of professionalization and externalization within a programme and leads to the emergence of explicit evaluation forms.

Uses of the APEP Grid

The APEP framework and grid are an attempt to develop evaluation theory. To theorize evaluation brings advantages and disadvantages for practitioners and researchers, and their expectations.

To theorize is to propose general laws, to build 'absolute' and not 'relative' knowledge, independent of specific context. Theory does not help a practitioner to solve specific problems. Is it useless? No, it is not, if the theory helps to come up with an answer. But it is the practitioner who must draw the practical implications from a theory.

To theorize is not to accumulate facts, but to generate hypotheses and to judge their feasibility. A theory is not generated from observed evaluation programme activities and outcomes, but verified by them. A theory can guide an evaluator or researcher but is only one step in a theorization process qualified by data.

Uses for Practitioners and Researchers

The APEP grid postulates the existence of an underlying commonality among various programme-evaluation practices, which vary depending on the context of evaluation use.

First, we shall attempt to show that the grid offers such a common framework for analysing the variety of programme-evaluation paradigms, models and practices. The first use is taxonomical. The APEP grid can help practitioners to choose a model suitable to a context.

Then, we shall discuss the grid's potential to train and advise practitioners to move towards more explication and formalization, and sometimes more externalization. We shall take an example from the field of health promotion.

A Classification

For each of the evaluation types described by the grid, the model that seems to correspond will be identified, based on literature.

Models and paradigms of evaluation stem from works and reviews by Alkin and Ellett (1990), House (1978, 1986), Monnier (1987), Scheerens (1990), Stufflebeam and Webster (1980), Stufflebeam and Shinkfield (1985), Worthen (1990).

T1: 'Decision' Evaluation (Explicit Decision) The 'decision' type of evaluation practice is distinguished along strategic and technical dimensions, as outlined above.

'Decision' evaluation encompasses what is often the first manifestation of evaluation present in internal evaluation practices. The decisions in those settings concern the daily, direct and immediate management activities of the programme. They are observable and explicit, but the enunciation of information processes and value judgements guiding these decisions remains largely implicit. The 'social need' of enunciation appears to be weak.

'Decision' evaluation can also address external decisions about the programme as well as political or administrative decisions. In that case, the social pressure is stronger for making the evaluation components more explicit. Both internal and external 'decision' evaluations become formal if decisions are taken formally (see Box 1).

Other forms of evaluation have to be identified against the background of the 'decision' type of evaluation, which is what makes this form of such interest.

T2: 'Expert' Evaluation (Explicit Judgement) No methodological procedures for information collection are used and the decision remains implicit. There are external and internal judgements, which may be either formal or informal.

Such external value judgements occur frequently and are particularly well organized socially. The core expertise is making the prototype of evaluation based on value judgements. This is in many ways the prototype for various forms of 'expert evaluation'. This type of evaluation is frequently encountered in audits, in the evaluation of scientific activities (within the researcher community) or in accreditation processes. Most evaluation taxonomies include it. Several authors

Box 1. Four Modes for 'Decision' Evaluation (Described by the APEP Grid)

'Decision' Evaluation Modes:

Form 1: informal internal explicit Decision	(iiD)
Form 2: formal internal explicit Decision	(fiD)
Form 3: informal external explicit Decision	(ieD)
Form 4: formal external explicit Decision	(feD)

(Alkin and Ellett, 1990; Monnier, 1987; Worthen, 1990) integrate professional expertise in their classification of the different evaluation forms.

Here we can also mention evaluation models centred on judicial procedures such as the consensus conference or parliamentary inquiry or court (court paradigm for Monnier, contradictory approach in the classification of House). Truth would emerge from the confrontation of viewpoints, the statement of opposed views.

According to Monnier, in the expertise paradigm, truth is based on the conviction of a person (or a group) known for his/her knowledge and competence. In such cases, the 'expert evaluation' is often informal. Judgement is formal if the criteria, norms, values on which the judgement bears are defined in advance.

Explicit 'expert' evaluation led by an internal actor is also common. This 'spontaneous' value judgement is often not well regarded by evaluation professionals and decision makers: internal experts have less 'social value'. But such experts can be useful and even 'objective', particularly when they operate formally making the basis for their judgement clear.

T3: 'Scientific' Evaluation (Explicit Process) For many evaluation specialists, the main goal of evaluation is to produce knowledge through an explicit formal and external process.

Evaluation specialists advocate 'scientific' methodologies for the collecting of information. The first paradigm of formal or 'scientific' evaluation is probably statistical (Monnier, 1987). The information process consists first in accumulating data and organizing them. The second paradigm is the experimental paradigm (Monnier, 1987). The information collected has to comply with criteria of 'truth' that only an experimental procedure would guarantee. The experimental paradigm in evaluation is the most frequently requested because it possesses the strongest credibility ('social value'). The third paradigm of the formal evaluation is the 'endosystemic' paradigm (Monnier, 1987). Methods here are 'naturalistic'. All three paradigms emphasize the methodological component of evaluation. In this professional evaluation approach, evaluation is deeply rooted in formal information systems and processes.

However, rather than only examine the methodological quality of an evaluation by reference to the experimental paradigm, it seems as useful to describe whether enunciation and a formal process are present or not. Moreover, we hypothesize that these two indicators (enunciation and formalization) will be more significant of the role of evaluation practice in the programme cycle than the choice of a given methodology for the process (design, quantitative or qualitative data, etc.). It is therefore interesting to see whether the programme manager signals the existence of an evaluation information process (enunciation) and makes its content more formal and precise (through indicators, tools, statistical analysis, etc.).

Now, programme managers can also collect information with an explicit internal process. This process can also be formal. As information producers, they are more likely to take on the responsibility of making use of acquired knowledge to become more efficient and to improve their practice.

Thus, to look at evaluation from the single angle of information-collection

processes is incomplete and partial. Evaluation is a continuous phenomenon. Whether internal or external, informal or formal, the explicit information process does not adequately describe the evaluation process.

T4: 'Control' Evaluation (Explicit Judgement and Decision) In this evaluation type, the information process remains implicit. The enunciation only of judgement and decision components signifies the social function attributed to this kind of evaluation. Various control evaluation practices can exist: they may be totally internal or mixed (either one or the other) and totally informal or mixed.

The administrative inspection is an example of a completely external and formal 'control' evaluation. It certainly is an evaluation practice, in that its aim is to give a value to an action in terms of established norms (most often limited to the verification of the respect of formal judicial rules) in order to come to an explicit and formal decision. It is a socially 'instituted' practice, because judgements (norms) and decisions (most often sanctions) are stated, observable and formal. But the external 'inspector' can use less formal evaluation practice. They can base themselves on an authority or a competence ('expert' evaluation) and take an explicit, but informally based decision.

'Control' evaluation can also be internal (with explicit judgement and decision), for example in an internal control of quality where only the information process remains implicit. The quality process can also be formal or informal. Lastly we can have different actors, internal and external: an external expert appraisal (judgement) leading to an internal decision, or an internal judgement leading to an external decision. These modalities are not evaluation models, but constitute different practices within the overall 'control' type of evaluation.

T5: 'Conclusion' Evaluation (Explicit Process and Judgement) The development of evaluation models centred on analysing information but aimed explicitly at expressing a value judgement, truly took off in the mid 1960s with authors like Scriven and Stake (Stufflebeam and Shinkfield, 1985). Evaluation here is oriented towards conclusions. Evaluation is defined as the systematic and objective determination of the value or the merit of an object or an action. Evaluation is comparative, undertaken professionally by an external and independent evaluator using systematic procedures. It results in judgements using 'objective' and explicit criteria.

Differences between models relate to methodology and the way information is analysed or the role of the evaluator as judge. (Scriven's model and Stake's model) share the same general form (as in Type 5, Figure 1): the importance both of the process of evaluation and of the value judgement that it allows. They are formal and external (process and judgement). Other 'conclusion' evaluation practices are possible with regard to strategic and technical dimensions:

- a completely internal and formal information process and judgement by a professional evaluator inserted in the management team;
- a collaboration between an external 'scientific' evaluator and an internal 'expert';

- an explicit internal and informal information process analysed for conclusions by an external 'expert'; and so on.

T6: 'Process-Decision' Evaluation (Explicit Process and Decision) The external and formal type also occurs frequently. Both the 'process' and 'decision' components are made explicit by the presence and the activities of the external professional evaluator and the external decision maker. The 'judgement' component is present, but often implicit. Those who have the legitimacy to decide also have, implicitly, the legitimacy to judge. In fact, the enunciation of the value judgement (produced by the evaluation process) will depend on the context of the production and the use of evaluation. In some contexts, the evaluation process may be totally explicit as in the case of a professional evaluator using the 'conclusion' evaluation or the decision maker using the 'control' evaluation.

Within the same diagram with two external actors, evaluation models differ depending on the nature and the degree of interaction between the evaluators and users. In a first model (Stufflebeam's model), evaluators delimit their work with the decision maker, then collect and provide information. The decision model is of a rational linear nature. In further models (Alkin, Patton), the decision making is seen as a more complex, incremental process where the evaluator constantly interacts with users (Scheerens, 1990).

Evaluator and decision maker can be internal, external or mixed – belonging to the management team or not. Process and decision can be informal or formal or mixed. The manager may report only informal information to an administrative decision maker. An external and formal 'scientific' evaluation may be examined by an internal board of administrators, and so on.

T7: 'Explicit' Evaluation (Explicit Process, Judgement and Decision) Explicit evaluation is the outcome of most evaluation practices that are current. Two examples show this development: the 'mediator' evaluation and the 'training' evaluation.

For Monnier (1987) the person in charge of evaluation can no longer be a traditional evaluator, i.e. simply an expert in methods and judge of programme value. He/she becomes a mediator. The sociology of organizations provides a better understanding of the management and evaluation of social problems. It involves a pluralistic process, never completed, with negotiations and compromises between protagonists. In this framework, the evaluator is involved in the political process. He or she is committed as mediator to be explicit and reach compromises on the definition of evaluation and action aims, on important information to collect, on the values at stake, on decisions to take. This person also has to discover and formulate proposals that mobilize widespread approval for follow-up action. The evaluator, according to our grid of analysis, becomes a co-decision maker. All of the components are explicit. They are external too. The 'mediator' evaluation model places centre stage an external evaluation role and person.

Explicit evaluation helps us better understand the close bond between action and evaluation processes, between the moment one gets information until a decision. For Vial (1987), what is at stake in education evaluation is not external

Evaluation 8(4)

control (through summative evaluation) or external support (through formative evaluation) of the learning process, but mastery of this process by learners themselves. 'Training' evaluation aims to provide the learner with the means to clarify and formalize what information to collect, the necessary criteria to evaluate and how to take decisions. This 'educational' vision of evaluation can be transposed to the evaluation programme. To view the programme manager or even the decision maker as a learner, can widen and enrich the self-evaluation concept that has been depreciated by many professional evaluators. Explicit evaluation can be internal too. Each component can also be informal or formal.

From T0 to T7

The totally implicit or totally explicit evaluations are the two extremes of a continuum. It is noteworthy that the eight types of evaluation allow us to identify and to give a name to evaluation practices at a given moment of their evolution. These are facets of the same phenomenon. Practices identified in a category can shift into another depending on the context and actors. All evaluation forms or models presented, represent intermediate positions on this continuum. An implicit management evaluation, an external expertise, a 'scientific' evaluation, an accreditation evaluation, an evaluation oriented towards assessing programme value, a decision-focused evaluation, a 'mediator' or 'training' evaluation, all can become at one moment or another totally implicit or explicit according to circumstances that do or do not necessitate enunciation.

In addition, if making explicit the components of evaluation offers a framework for classifying evaluation models, it is also necessary to use the technical and strategic axes to describe the practices that have evolved within the different types of evaluation. For example, the position of the different actors and their 'specialization', allows us to understand the apparent discontinuity of evaluation activity. This division of labour is all the greater as programmes become larger and involve more people.

The difficulty of situating a practice precisely indicates the limits of a theoretical approach. The theoretical framework and the grid are only tools to enable the practitioner to choose a model suitable to a specific situation on the field.

A Quality Improvement Process

As well as allowing a classification of the main evaluation forms, the APEP theoretical framework also helps analyse the underlying issues in evaluation practice. It does so by using the three axes of differentiation: social (enunciation), strategic (externalization) and technical (formalization).

The first example shows the issues at stake in health service accreditation. The next example attempts to show how these three axes can be used to improve an expert evaluation (Lafontaine et al., 2001).

The Accreditation of Health Services

In the health service evaluation field, the APEP grid can describe different 'expert' evaluation models (T2) developed on the strategic axis (internal-external) and the technical axis (informal-formal).

- Internal development ('self-improvement') is based on an **informal** and **internal Judgement**. (**iiJ**)
- The examination by peers ('peer review') is based on an **informal** and **external Judgement**. (**ieJ**)
- Recommended professional norms ('guidelines') lead to a **formal** and **internal** or **external Judgement** depending on the origin of the standardized criteria. (**fiJ** or **feJ**)
- Accreditation by an independent body ('accreditation') is a **formal** and **external Judgement**. (**feJ**)

The 'ideal' model (accreditation) is characterized by norms or standards measured by external and independent evaluators. Therefore, it is necessary to distinguish two distinct – although connected – concepts in accreditation: quality and accountability. Quality implies a constant search for the improvement of services to attain ever-higher standards. Accountability requires the public proof of adequate security levels, efficacy and efficiency.

Scrivens (1997), who has analysed systems of health service accreditation in the USA, Canada, Australia and the United Kingdom, suggests that highly externalized systems and highly formalized ones lead to weaker internal utilization. Accreditation systems protect the public and reassure financial sponsors. On the other hand they reduce the professional development based on the exchange (peer review) or the internal improvement (self-improvement) model.

The issue at stake is how to give a 'guarantee of quality' whilst encouraging continuous improvement. One way is to strengthen the 'social value' of internal evaluation by means of its enunciation and formalization. Quality certification that also attests the quality of internal evaluation works in this direction.

The Expert Evaluation of Health Promotion Programmes or Material

In certain contexts, it is not possible to use evidence of efficiency to evaluate the quality of a programme or material.

Thus, for example:

- the decision whether to subsidize a project is generally taken on the basis of a written document before the implementation of the programme;
- the promotion of educational material (e.g. by presenting it in a catalogue or on a website) is usually not based on a rigorous testing of this educational material in different contexts of use.

One can then resort to expert evaluation to assess the quality of the projects or materials in question.

In expert evaluation, as it is usually defined, the judgement components are assigned to external expert bodies (that are neither promoters of the programme, nor producers of the material). The only requirement is that these experts explicitly formulate a judgement on the programme or the material. Information is available to the expert, but it is rarely explicit and formalized.

The decision is not the responsibility of experts. Depending on the case, it may be the responsibility of the budgetary authority, of the people who are responsible for the publishing of catalogues and websites or of users who wish to select material.

Since 1984, the team of the APES-ULG (University of Liège) has been working towards the improvement of the devices of 'expert' evaluation, in particular by developing methods and materials that strengthen the clarification and the formalization of the following three components: process, judgement and decision.

Ways of Clarifying the Process and Judgement Components

The process is explicit since the information used by experts to pass their judgement is identified, as is the procedure to produce the judgement. The judgement is explicit since a report is written and presented by experts.

Formalizing the Process and Judgement Components This has been at the centre of the scientific and methodological support provided by the APES-ULG.

Formalizing the Judgement Component Criteria have been defined and formalized in analysis grids to judge the quality:

- of a programme, i.e. its priority, coherence, pertinence, usefulness, credibility;
- of an educational support, i.e. accuracy and pertinence of the content, interactivity, attractiveness.

Formalizing the Process Component The sequencing and elaboration of expertise judgement are formally planned through:

- the presentation of the programme according to a standardized model, i.e. perusal by two experts who have to write a report, discussion by an advisory 'commission', writing of an experts' report based on consensus;
- the ways experts explore the material, individual use of the analysis grids, collective writing of a report based on a synthesis form that brings together different analyses.

The progressive adjustments of these two aspects of expert evaluation (by formalizing judgement and the process) in the French Community of Belgium lead us to conclude that analysis grids and the collective write-up of expert judgements are two requirements that will increase the quality of expertise-based evaluation outputs.

A third quality requirement has been particularly underlined by the way experts explore material. Experts need to be encouraged and trained to put into practice standardized analysis procedures when producing reports.

A last issue is linked to the willingness to use the materials and products of evaluation expertise to improve the practice of programme promoters and material 'creators', on the one hand, and decision makers and users of materials, on the other.

One observes that the appropriation of these recommendations by producers and promoters of projects causes problems: the results of expert evaluation remain external, as was the process of evaluation.

Indeed, promoters, producers, decision makers and users do not currently help define the criteria of judgement. Promoters of programmes participate in the first stage of the process of evaluation by agreeing to write the project according to a plan (a writing model) that facilitates the use of judgement (evaluation criteria) by experts. The creators of material, on the other hand, are not asked to make a standardized presentation of their products for the committee of experts.

At the current stage, the new challenge that we, experts (inventors of the process and analysis grids), have to take up is to think critically about this relative absence of the producer and the decision maker in the process of expert evaluation. To clarify and formalize decision components is the new issue to be addressed in order to move forward on 'explicit' evaluation (T7).

Uses for Researchers

Between the theory and the reality of what is at stake – the concrete situation on the field – there unfolds a distance that is necessary to be explored. To identify the diversity of the programme-evaluation practices, to describe it (how), to understand it (why), and later, to be able to exploit it at its best and to reply to the needs of evaluation actors, such would be the aim of research on the programme-evaluation practice.

We have identified briefly three attempts to describe programme-evaluation practices: taxonomies, research on models and research on practices. The APEP grid is, in a way, a taxonomy but with underlying dimensions that ought to allow research on evaluation practice.

The theory framework and the APEP grid tool can guide the researcher. Where to observe? Which practices to identify?

We put them to the test on two occasions, which we shall only present briefly in this article: to identify and quantify internal evaluation practices; and to study the relationship between professional empowerment and evaluation practices.

Firstly, we made a retrospective study of programme-evaluation practices in 46 health education programme project funding applications managed in the French Community of Belgium in 1992 and 1993 (Demarteau, 1998: 209). Working on funding applications brings both advantages and limits. A funding application is a written document, a public product, put forward in a particular context that is likely to make evident an enunciated evaluation practice. It especially facilitates the location and analysis of explicit internal evaluation practices. On the other hand, it does not yet address effective practices. The next step will be the study of activity reports and other public products with explicit evaluation practices, which will also bring other advantages and limits. The text content of funding applications has been systematically coded line by line.

We have studied the frequencies and conditions within each of the evaluation modes (i.e. any one of the 125 combinations predicted by the conceptual framework) in relation to the role of the programme evaluation in decision making and

Evaluation 8(4)

its objectives, i.e. health promotion strategies. This follows from the Stufflebeam CIPP model. We have also taken into account different contextual characteristics such as the targeted public, the evaluator's status, his/her experience and evaluation budgets.

Analysis confirms the great diversity of programme-evaluation practices. Nature and frequency of the evaluation mode are functions of the type of decision making (Context, Input, Process, Product), the diversity of evaluation objects and the professionalization of actors.

For example, Table 1 shows the frequency of some evaluation modes in 'product' evaluation. In this case, 'scientific' evaluation (T3) and 'conclusion' evaluation (T5) are most frequent. Internal practices are more frequent than external ones. Internal practices are more often formal, while external evaluations are more often informal. It is as if resorting to an external evaluator was a sufficient guarantee of quality. The 'guarantee' phenomenon is also obvious for the 'expert' evaluation (T2), and was even more common. 'Expert' evaluation in our study was always informal.

At the end of this study and within its limits, we can demonstrate convergence between the social context of a project, its content, its structure and its evaluation mode. Evaluations are inevitably first a social practice before they become methodological.

The APEP framework can also be used to study individual evaluation programmes. Evaluation components and their axes have been used to construct a descriptive 'evaluation practice scale'. This scale has been validated in an extensive study (Demarteau and Thiess, 2000) on the relationship between the perception of autonomy in the conduct of a health promotion programme (professional empowerment) and the use of a particular evaluation practice.

The aim of this study was to verify the hypothesis that there are different evaluation practices for differently perceived levels of professional empowerment

Table 1. Frequency of some Evaluation Forms in Product Evaluation (31 projects)

<i>n</i> = 31	<i>T1</i>	<i>T2</i>	<i>T3</i>	<i>T4</i>	<i>T5</i>	<i>T6</i>	<i>T7</i>
TOTAL	0	3	23	0	9	5	4
Internal	0	1	17	0	8	4	4
Informal	0	1 iiJ	6 iiP	0	5 iiP iiJ	1 iiP iiD	1 iiP iiJ iiD
Formal	0	0	11 fiP	0	3 fiP fiJ	3 fiP fiD	3 fiP fiJ fiD
External	0	2	6	0	1	1	0
Informal	0	2 ieJ	5 ieP	0	1 ieP ieJ	1 ieP ieD	0
Formal	0	0	1 feP	0	0	0	0

Note: formal–informal; external–internal; Process–Judgement–Decision.

and for other independent variables such as length of service, work time and specific training in health promotion. Two Likert scales about empowerment perceptions by health promotion professionals and about their evaluation practices have been designed and validated. Self-administered questionnaires were sent to 289 health promotion professionals active in French-speaking Belgium; 100 answers (34.6%) have been analysed.

The relation between professional empowerment perception and evaluation practices is significant. The description of professional practices and the need to understand them appear to be a preliminary step for the development and improvement of health promotion programme evaluation.

Conclusions

It is necessary to describe and understand programme evaluation practice in order to propose better practices and to contribute to developing quality programmes. Our initial goal was to elaborate a framework that would allow us to describe and explain all programme evaluation practices that are likely to be met in the context of publicly financed health education programmes.

This article does not claim to be primarily a work of research or practice; it is rather a theoretical work. It proposes an extensive framework to identify the initial forms of evaluation practice. This tool will thus be useless for the professional evaluator, but useful, we hope, for the project manager or decision/policy maker, for the trainer or the consultant in evaluation practice. It shows ways to improve practice by expressing, formalizing and externalizing evaluation components.

- The APEP grid allows us to integrate the variety of the programme evaluation paradigms and 'models' and provides an analysis framework for comparative purposes.
- It underlines the issues at stake in the relationship between knowledge, judgement and decision.

At the same time, we are continuing the analysis of these health education programmes by examining activity reports to compare evaluation proposals with effectively realized evaluations.

In addition we see the potential to test the APEP framework in other evaluation practice/contexts. One can hypothesize that different social conditions would invoke different enunciation modes. For example, do programmes financed by the European Union (in an international context) favour more external and formalized evaluation practices? If so, do they become a substitute for internal evaluation practices? Public health also provides other opportunities to analyse the influence of contexts on evaluation practices – for example, community health programmes and health personnel training programmes.

Note

1. Author's own translation of the original French text.

References

- Alkin, M. C. and F. S. Ellett (1979) 'An Inquiry into the Nature of Evaluation Theorizing', *Studies in Educational Evaluation* 5(2): 151-6.
- Alkin, M. C. and F. S. Ellett (1990) 'Development of Evaluation Models', in H. J. Walberg and G. D. Haertel (eds) *The International Encyclopedia of Educational Evaluation*, pp. 15-21. London: Pergamon Press.
- Barbier, J. M. (1990) *L'évaluation en Formation*. Paris: PUF.
- Cousins, J. B. and K. A. Leithwood (1986) 'Current Empirical Research on Evaluation Utilization', *Review of Educational Research* 56(3): 331-64.
- Crozier, M. and E. Friedberg (1977) *L'acteur et le système*. Paris: Seuil.
- Demarteau, M. (1994) 'Le conseil en évaluation: une fonction professionnelle à développer pour une évolution des pratiques de promotion de la santé', Paper presented at the *Troisièmes Rencontres Francophones Internationales de Promotion de la Santé*. Lille, 27-30 June.
- Demarteau, M. (1998) *Analyse des pratiques d'évaluation de programmes d'éducation pour la santé*, (Thesis in Public Health) University of Liège.
- Demarteau, M. and M. Thiess (2000) 'Professional Empowerment and Evaluation Practices in Health Promotion', Paper presented at the *3rd Nordic Health Promotion Research Conference 'Outcomes in Health Promotion'*. September 6-9, Tampere, Finland.
- Guba, E. G. and Y. S. Lincoln (1989) *Fourth Generation Evaluation*. London: Sage.
- House, E. R. (1978) 'Assumptions Underlying Evaluation Models', *Educational Researcher* 8: 4-12.
- House, E. R. (ed.) (1986) *New Directions in Educational Evaluation*. London: Falmer Press.
- Lafontaine, C., C. Vandoorne and M. Demarteau (2001) 'Issues at Stake and Requirements of a Quality Improvement of Expert Evaluation', Paper presented at the *XVIIth World IUHPE Conference*, 15-20 July, Paris, France.
- Marsch, D. D., W. B. Newman and W. F. Bover (1981) *Comparing Ideal and Real. A Study of Evaluation Practice in California Using the Joint Committee's Evaluation Standards*. Los Angeles, CA: AERA.
- Monnier, E. (1987) *Evaluations de l'action des pouvoirs publics: du projet au bilan*. Paris: Economica.
- Nevo, D. (1986) 'The Conceptualization of Educational Evaluation: an Analytical Review of the Literature', in E. R. House (ed.) *New Directions in Educational Evaluation*, pp. 15-29. Philadelphia, PA: Falmer Press.
- Patton, M. (1978) *Utilisation-focused Evaluation*. London: Sage.
- Scheerens, J. (1990) 'Beyond Decision-oriented Evaluation', in H. J. Walberg and G. D. Haertel (eds) *The International Encyclopedia of Educational Evaluation*, pp. 35-40. London: Pergamon Press.
- Scrivens, E. (1997) 'Assessing the Value of Accreditation Systems', *European Journal of Public Health* 7(1): 4-8.
- Shadish, W. R. and R. Epstein (1987) 'Patterns of Program Evaluation Practice among Members of the Evaluation Research Society and Evaluation Network', *Evaluation Review* 11: 555-90.
- Stufflebeam, D. L. and A. J. Shinkfield (1985) *Systematic Evaluation: A Self-instructional Guide to Theory and Practice*. Boston, MA: Kluwer-Nijhoff.
- Stufflebeam, D. L. and W. J. Webster (1980) 'An Analysis of Alternative Approaches to Evaluation', *Educational Evaluation and Policy Analysis* 2(3): 5-20.

Demarteau: A Theoretical Framework and Grid for APEP

- Vial, M. P. (1987) 'Evaluer n'est pas mesurer', *Cahiers Pédagogiques* 16–17.
- Williams, J. E. (1989) 'A Numerically Developed Taxonomy of Evaluation Theory and Practice', *Evaluation Review* 13(1): 18–31.
- Worthen, B. R. (1990) 'Program Evaluation', in H. J. Walberg and G. D. Haertel (eds) *The International Encyclopedia of Educational Evaluation*, pp. 42–7. London: Pergamon Press.

MICHEL DEMARTEAU is head of the department of Health Promotion and Prevention at the Public Health Institute of *Province de Hainaut*. He is responsible for implementing health promotion and disease prevention programmes. Please address correspondence to: Observatoire de la Santé du Hainaut, Rue Saint-Antoine 1, B-7021 Havré, Belgium. [email: michel.demarteau@hainaut.be]