Choosing Evaluation Models

A Discussion on Evaluation Design

HANNE FOSS HANSEN
University of Copenhagen, Denmark

A variety of different evaluation models are presented in the evaluation literature. These mostly fall into the following categories: results models, process models, system models, economic models, actor models, and programme theory models. This raises the question: 'how can evaluation sponsors and evaluators decide how to design an evaluation with so many models to choose from?' In this article, several – mutually incompatible – recommendations are discussed. Design should be determined by the purpose of the evaluation, the object of evaluation or the problem to be solved by the evaluated programme or agency. The recommendations are based on different rationales: goals–means, context-based values and programme theory. Furthermore, in practice other logics may influence the evaluation design processes. Four hypotheses concerning such logics are proposed: negotiation, appropriateness, 'routine' and projection of competence.

KEYWORDS: designing evaluation; evaluation models; logics of design; programme theory of evaluation

Introduction

The evaluation literature offers a rich variety of alternative approaches to evaluation. Over the years, the evaluator's toolbox has continuously expanded. The options are multiple, such as opting for summative or formative evaluation or stressing the clients', experts' or general stakeholders' concerns. But the choice is also between subscribing to realistic evaluation, theory-based evaluation, utilization-focused evaluation or empowerment evaluation, just to mention a few examples. Authors of much of the literature on evaluation claim their particular evaluation model to be the best. However, we need to study the alternative approaches to evaluation comparatively and to develop meta-models in order to enable us to reflect upon designing evaluations.

The wide variety of alternative approaches in the evaluator's toolbox raises the important question of what criteria should be used to compare one approach with another or perhaps decide to combine several approaches. This issue will be discussed in this article.

447

First the concept of evaluation will be discussed. Second, a typology of generic evaluation approaches, i.e. evaluation models, is presented. Third, different criteria for choosing models and designing evaluation are discussed. Finally, reflections on other factors that might influence evaluation design are presented.

The Concept of Evaluation

The concept of evaluation can be defined as 'a study designed and conducted to assist some audience to assess an object's merit and worth' (Stufflebeam, 2000: 35); or in the same vein as a 'careful retrospective assessment of the merit, worth and value of administration, output and outcome of government interventions, which is intended to play a role in future, practical action situations' (Vedung, 1997: 3). Both definitions seem to provide for two theoretical traditions, which academic literature, strangely enough, has addressed relatively separately. The two traditions are programme evaluation (Stufflebeam, 2000, offers a good outline of approaches within this field) and organization evaluation; the latter is most often referred to as approaches to assess organizational effectiveness (Scott, 2003, offers a good overview of approaches in this field).

As the term programme evaluation suggests, the body of literature in this field focuses on assessments of programmes, defined as, for example, an 'organized, planned, and usually ongoing effort designed to ameliorate a social problem or improve social conditions' (Rossi et al., 2004: 29). The literature on organizational effectiveness focuses on the efforts of organizations and is somewhat generic in the sense that the intention has been to direct it towards, and be relevant for, all types of organizations, whether they are private or public, big or small, characterized by standardized or flexible production, etc. (Cameron, 1986; Scott, 2003). In recent years the focus has shifted towards how the generic tradition can be adapted and further developed in order to be of greater relevance for understanding the conception of effectiveness characterizing public organizations (Hansen, 1999).

It is relevant to include the two academic traditions in a discussion about evaluation design for several reasons. First, both traditions basically deal with how to conduct assessments, including on which approaches, criteria and values to base the assessments. Second, the programme evaluation literature is increasingly recognizing that its evaluation models are also applicable to organizations (e.g. Owen and Rogers, 1999: 33; Fetterman, 1996: 4). Finally, the basic evaluation models employed within the two traditions overlap significantly, as will be demonstrated in the next section.

A Typology of Evaluation Models

An evaluation model stipulates the question that a given type of evaluation seeks to answer, as well as specifies how to set up the criteria for assessment. Both the literature on programme evaluation and that on organizational effectiveness offer several typologies of evaluation models. The typology presented in Table 1 primarily draws inspiration from Cameron (1986), Hansen (1989, 2003), Scriven (2003) and Vedung (1997).

448

Table 1. A Typology of Evaluation Models

Evaluation Models	Questions	Criteria for Evaluation
Result models		
a) Goal-attainment model	a) To what degree has the goal(s) been realized?	a) Derived from goal(s)
b) Effects model	b) Which effects can be uncovered?	b) Open, all consequences should be uncovered
Explanatory process model	Is the level of activity satisfactory? Are there implementation problems?	Performance is analysed from idea to decision and implementation and to the reaction of the addressees
System model	How has performance functioned as a whole?	Realized input, process, structure and outcome assessed either in relation to objectives in same dimensions or comparatively
Economic model		
a) Cost-efficiency	a) Is productivity satisfactory?	a) Output measured in relation to expenses
b) Cost-effectiveness	b) Is effectiveness satisfactory?	b) Effect measured in relation to expenses
c) Cost-benefit	c) Is utility satisfactory?	c) Utility measured in relation to expenses
Actor model		
a) Client-oriented model	a) Are clients satisfied?	a) Formulated by clients
b) Stakeholder model	b) Are stakeholders satisfied?	b) Formulated by stakeholders
c) Peer review model	c) Is professional quality in order?	c) Formulated by peers
Programme theory model (theory-based evaluation)	What works for whom in which context? Is it possible to ascertain errors in programme theory?	Programme theory is reconstructed and assessed via empirical analysis

In general, evaluation models fall into six categories: results models, process models, system models, economic models, actor models and programme theory models.

The *results models*, or summative evaluation, focus on the results of a given performance, programme or organization. There are several subcategories of results models. In the goal-attainment model, which is the classic model in the literature on programme evaluation and organizational evaluation, results are assessed only in relation to the predetermined goals. The evaluator closely scrutinizes the object of evaluation in relation to its goals.

449

The *effects model* involves a broader scope. The intention here is to elucidate all the consequences of the object that is subject to evaluation. The effects model, sometimes referred to as the goal-free evaluation model (Scriven, 1973), has been criticized for deficient evaluative criteria. The evaluator may circumvent this problem by applying the model to a before/after analysis, i.e. to analyse the organizational field before and after the given performance, programme or organization was implemented.

Process models focus, naturally, on the ongoing processes and efforts. The explanatory process evaluation follows an intervention from its early phase of being merely an idea to the decision to implement it, its implementation and reception among clients and other stakeholders. Explanatory process evaluation is ideally executed 'real-time', less ideally via historical analyses 'unravelling' the process 'chains'.

The overall orientation of the *system model* is a system perspective, which analyses input, structure, process and outcome in terms of results. The assessment can be based either on comparisons of planned and realized input, structure, process and results or on certain forms of benchmarking that compare the results with similar programmes or organizations that are considered excellent.

The *economic models* to some extent also build on a system perspective. However, these models differ somewhat from ordinary system models in that they consider the object of evaluation – the programme or the organization – as a 'black box' by relating the assessment of results (either performance in the form of output, effects or more lasting benefits) to the expenses involved (input).

Actor models are based upon the actors' own criteria for assessment. The client-oriented model focuses on the clients' criteria for assessment, the stake-holder model on all the relevant interested parties' criteria for assessment, and the peer review model on the criteria for assessment of professionals, the most outstanding members of the profession.

Finally, the *programme theory model* focuses on assessing the validity of the programme theory on which the given intervention or organization builds. Programme theory is either reconstructed and compared with empirical analysis of problems and problem-solving (as suggested in the 'theory-based' evaluation approach by Birckmayer and Weiss, 2000), or empirical observation is used to analyse the causal relations between context, mechanism and outcome (as suggested in the 'realistic evaluation' approach by Pawson and Tilley, 1997 and expanded in the 'realist synthesis' approach by Pawson, 2002). The programme theory model may be seen as an extended results model. Where the classical results models focus only on concrete goals or effects related to a specific inter-vention or organization, and aims to evaluate the results of these, the programme theory model opens up the underlying black box of the programme theory, uncovers mechanisms and raises the focus to a cluster of interventions or to an organizational field. The aim of the programme theory model is to revise and further develop programme theory and thus learn what works for whom in which contexts.

Reviewing the historical development of the field of evaluation, it is apparent that an increasing number of evaluation models are developed over time. Some

450

of the models, in particular the goal-attainment model and the economic models, build on a notion of objectivity. They have a built-in 'assessment optimism' in the sense that they build on the notion that evaluation can yield objective results. Other models, the stakeholder model in particular, challenge this perspective. The stakeholder model includes a conflict perspective. The criteria for assessment are not clearly stated ex ante, but formulated in the evaluation process.

From a technical, methodological perspective, all models have their strengths and weaknesses, which due to space will not be discussed here in detail. However, the issue can be illustrated with the help of photography. The choice of a model corresponds to zooming in and taking a picture. The choice of a model is the choice of a field of vision. The choice of a model (or combination of models) thus entails that certain aspects fall into focus, while others are excluded.

The current situation is that of multiple competing models. From a theoretical, analytical perspective this situation is interesting, but from a practical position the multitude of evaluation models available seems overwhelming and confusing. How are evaluation sponsors and/or evaluators to choose from the heterogeneous range of evaluation models in the toolbox? One of the problems is that the variety of evaluation approaches and evaluation models has not been matched by a corresponding increase in thinking about the choice of model(s). Considering the literature on evaluation and organizational effectiveness, it is striking that there are so few theoretical, methodological discussions about principles and criteria for the choice of (combinations of) evaluation models.

However, three different types of logic, identified in the literature, are discussed here. First, that evaluation design may be based on the purpose of carrying out an evaluation; second, on the characteristics of the evaluand; and third, on the characteristics of the problem that the programme or organization under evaluation aims to resolve.

Criterion I: The Purpose of Evaluation should Determine Design

The first chain of reasoning is based on the purpose of the evaluation, arguing that purpose should determine the choice of evaluation model. The literature on programme evaluation (Rossi et al., 2004: 34; Scriven, 1991) recommends formative evaluation if the evaluation is intended to create learning and thereby improve the programme in question. Correspondingly, summative evaluation is recommended if the evaluation is intended to control performance in terms of accountability. This reasoning is developed in greatest detail by Premfors (1989), whose recommendations are illustrated in Table 2.

As the table illustrates, it is recommended that the purpose of the evaluation determine the organizing of the evaluation, the collection of data, and the dissemination and use of its results. If the purpose is control, the recommendation is to base the design of the evaluation on results models and in particular on the goal-attainment model. If the purpose is learning, it is recommended that the design of the evaluation is based on the stakeholder model. On the face of it, this goals–means reasoning seems plausible, but is the best advice always to let the

451

Table 2. Choice of Evaluation Model on the Basis of the Purpose of the Evaluation

Evaluation Design	Evaluation Purpose	
	Control	Learning
Evaluation model	Primary emphasis on result models	Primary emphasis on process models
Organization	Top-down	Self-evaluation or independent peers/consultants
Criteria	Derived from goals on higher level	Derived from stakeholder requests
Method	Primary emphasis on quantitative measurement of effects	Primary emphasis on qualitative process studies
Dissemination of results	Upwards in hierarchy	Via interactive process to broad groups of interested parties
Utilization	Foundation for sanctions and control on higher level	Foundation for self-conception and local organizational development

purpose of an evaluation determine the choice of (combinations of) evaluation models? The goals–means logic and the recommendation to opt for results models if the purpose of evaluation is to control implicitly presuppose two premises. The first premise is the existence of reliable methods for results measurement. The other is that those responsible for controlling have insight into the world of those subject to evaluation. The absence of such insight renders difficult interpretation of results measurement. Obviously the potential risk is that erroneous interpretations of evaluation results create situations in which controlling and sanctioning may do more harm than good.

Viewed in the light of these facts, the questions are as follows. How many policy areas have developed reliable methods for results measurement? And within how many policy areas does the responsible political-administrative authority have sufficient insight into relations on the 'producing' level? Providing an answer to these questions naturally requires extensive discussions of individual policy areas, as well as of the concrete methods developed for results measurement. For reasons of space this will not be discussed here, but it is obvious that quite a number of the complex policy areas within which evaluation is conducted in practice do not meet the two conditions. The area of research policy is an example. Here many analyses show that results measurement, which in this context builds on bibliometric and scientometric methods, should be used with considerable caution, as the validity of the methods is highly limited (Hansen and Jørgensen, 1995; Seglen, 1992, 1994).

Finally, it could be argued that the recommendation to design evaluations according to purpose rests on the premise of a clear purpose having been formulated a priori. Once again you could ask how often is the purpose of an evaluation clearly formulated. And what are the implications if the purpose is ambiguous, or as often is the case in practice, the purpose includes aspects of both

452

control and learning? In any case, in this light you might conclude that there are several good reasons for discussing other criteria and principles than the goals–means rationale as the premise for the choice of (combinations of) evaluation models.

Criterion II: Characteristics of the Evaluand should Determine Design

Another set of recommendations has an entirely different basis. The reasoning here is that the characteristics of the object to be evaluated, the evaluand, must determine the choice of (combinations of) evaluation model(s). This reasoning runs parallel with the contingency perspective in organization theory, which states that characteristics of the tasks and the environment of an organization are the premises for structural design (e.g. Mintzberg, 1983).

There are two variants of this reasoning: (1) choice of model(s) should be determined by what is possible in view of the characteristics of the evaluand (possibility reasoning); (2) choice of model(s) should be determined in view of what the characteristics of the evaluand justify (legitimacy and justice reasoning).

Possibility Reasoning

Possibility reasoning is based on the view that some evaluation models are stronger than others and therefore preferable to use. It is argued that it is optimal to evaluate on the basis of the economic models. If this is not possible, goal-attainment evaluation should be used. If neither is possible, entirely different paths must be pursued.

Table 3 demonstrates that two aspects of the evaluand determine which evaluation model it is possible to use. These aspects are the purpose of the object and knowledge of the programme theory on which the efforts build.

Following this reasoning it is only possible to select an economic model if the objectives of the evaluated object (a programme or an organization) are clear and if knowledge of the programme theory on which activities build is complete. Only under these conditions can an economic model evaluate whether a performance has been successfully optimized. If the objective is clear, but knowledge about the programme theory incomplete, the strongest evaluation design available is the goal-attainment model. In other words, the goal-attainment model can be chosen only if goals are clear.

Table 3. Choice of Evaluation Model: Possibility Reasoning (freely from Thompson, 1967: 86)

Objective	Knowledge about Programme Theory	
	Perfect	Imperfect
Clear	Economic models	Goal-attainment model
Unclear	Comparative evaluation	Comparative evaluation

453

In case objectives are unclear, the evaluation should be designed from a comparative perspective. An evaluation must be designed so that it either assesses the development of the object over time or compares it with a given standard or a reference group such as one or more related projects, organizations or programmes.

A comparative evaluation can be based on several alternative evaluation models. This perspective is well developed in the literature on benchmarking (e.g. Foss, 2000). The literature distinguishes between standard benchmarking (assessment against a given performance standard), results benchmarking (comparing the performance of a large number of public organizations), and process benchmarking (comparing an organization's production process with that of an organization considered to be leading in the field – 'best practice'). Summing up, possibility reasoning helps specify the conditions that must be present in order for an evaluation to be designed on the basis of economic models or the goal-attainment model. This reasoning might seem to be of limited use. However, if we reflect on how widespread the goal-attainment model is in practice, possibility reasoning raises the question of how often the conditions for choosing the goal-attainment model actually are fulfilled compared with the model's prevalence.

Cameron (1986) is an example of an author whose writing is rooted in possibility reasoning. He argues that using the goal-attainment model is only possible and useful if the goals of the evaluated object are clear, if there is consensus about them, and if they are time-bound and measurable. He argues further that using a process model is only relevant if there is a clear relationship between processes and results, and using a stakeholder model is only relevant if stakeholders are able to influence the programme or organization evaluated.

Legitimacy and Justice Reasoning
Another criterion and design principle that is tied to characteristics of the evaluated object can be labelled legitimacy and justice reasoning. Table 4 illustrates this reasoning.

The basis for legitimacy and justice reasoning is that public sectors are heterogeneous. Problem-solving is organized differently in different national contexts and in different policy fields. Public efforts, programmes and organizations are anchored in different forms of networks, values and contexts. Jørgensen (2003) distinguishes between four ideal-types of context, each representing a different model of the state.

Table 4. Choice of Evaluation Model: Legitimacy and Justice Reasoning (inspired by Jørgensen, 2003; Hansen, 2003)

State Model	Organizational Context	Evaluation Model
Sovereign state	Parliamentary chain of command	Goal attainment model
Autonomous state	Professions	Peer review model
Negotiated state	Interested parties	Stakeholder model
Responsive state	Clients	Client-oriented model

454

The sovereign state is tied to a conception of classical representative democracy and the parliamentary chain of control. Politicians are elected to formulate overriding objectives for the development of society and implement them through legislation and intervention. Agencies and public organizations are governed politically and are accountable for implementation. The equal treatment of citizens and corporations is central. The citizens' primary role is to elect the politicians.

In the autonomous state, on the other hand, the administration and public organizations are independent carriers of values, e.g. values related to health, environment, pedagogical values, etc. Citizens and corporations (and politicians, for that matter) are guided and socialized. A technocracy founded on professional networks and professional organizations influences activities and performances.

The negotiating state builds on a conception of real politics. Many interested parties, including political parties, commercial organizations, labour market organizations, professional associations and consumer groups, negotiate and influence the development. Agencies and public organizations are intermediaries and mediate among different interests.

Finally, the responsive state builds on a conception of the client's needs and demands for services as being central. The role of politicians is to ensure the responsiveness of agencies and public organizations. This responsiveness can be secured via several mechanisms, such as the employment of market mechanisms or user boards.

As already mentioned, these state models are ideal models that cannot be found in the 'real world', where various combinations prevail. At the same time, even though there are different national administrative styles, it is apparent that different policy areas are influenced by different models to varying degrees. Naturally the sovereign state model is recognizable to a certain extent in all policy fields, but in particular in rule-regulated areas, such as those managing regulatory tasks. Profession-dominated areas, such as health, education, research and culture, are influenced by the autonomous state model. The negotiating state model is to be found in particular in labour market and industrial policy areas, and the responsive state model in the major service areas, including traffic, children, elderly people.

Every state model is matched by an evaluation model. The sovereign state is matched by the goal-attainment model, which can furnish politicians with feedback as to the extent to which the formulated goals have been achieved. The autonomous state is matched by the peer review model, which can be applied to assess the extent to which the professional quality is satisfactory. The negotiated state is matched by the stakeholder evaluation model, which can provide feedback on relevant interested parties' assessment of performance and activity. Finally, the responsive state is matched by the client-oriented model, which can provide feedback on user needs and assessments.

The third recommendation regarding the choice of (combinations of) evaluation models is based on legitimacy and justice reasoning. The core of this reasoning is the following. Evaluation of policy areas influenced by the sovereign state

should be based on the goal-attainment model. Evaluations of policy areas influenced by the autonomous state should be based on the peer review model. Evaluations of policy areas influenced by the negotiated state should be based on the stakeholder model, and evaluations of policy areas influenced by the responsive state should be based on the client-oriented model.

Only evaluation designs that match the key characteristics and values of the evaluated areas will be considered as legitimate and just and therefore viable (Hansen and Borum, 1999, demonstrate that this has been the case in the university sector in Denmark).

This legitimacy and justice reasoning seems plausible. However, it does raise questions such as whether evaluations designed according to this logic will tend to have a 'conservation' effect, thus reinforcing activities that are consistent with the assumptions already characterizing a particular area.

Change Reasoning

This risk of conservatism raises the question of whether to use change reasoning rather than legitimacy and justice reasoning. If the purpose is, for example, to increase responsiveness towards clients in an area strongly influenced by the sovereign state, would it not be better to design evaluations based on the client-oriented model? Or, maybe using the goal-attainment model would strengthen 'the bureaucratic backbone' of areas that are, for example, influenced by the negotiated state? Change reasoning, which combines the recommendation that the object of evaluation should determine design with that of the purpose of evaluation should determine design, is illustrated in Table 5.

Accordingly, conflicts are bound to occur over which evaluation design will be perceived as legitimate and just and which will be perceived as appropriate, if the evaluation is used as part of a reform strategy.

Criterion III: The Problem to be Solved by the Evaluated Object should Determine Design

The third set of recommendations is rooted in the understanding that evaluation design should be determined on the basis of an analysis of the problem that the object of evaluation is meant to resolve. The potential of evaluations designed on the basis of this reasoning is to further develop the programme theory. As mentioned earlier, this approach is referred to as 'realist evaluation' and 'realist synthesis' as well as theory-based evaluation.

These approaches have proved fruitful in several policy fields (e.g. crime prevention and health promotion). They seem, however, difficult to apply to evaluations of very complex and 'integrated' interventions, such as public sector reforms and regulatory reforms in which several actors jointly and in networks try to tackle simultaneously various aspects of a problem (see also Stame, 2004). The aim to identify causality between context, mechanism and outcome is difficult to pursue when the number of variables increases. In other words, it becomes unrealistic to use the realist approach if it is impossible to specify all important variables.

456

Table 5. Choice of Evaluation Model: A Strategy for Change (Hansen, 2003)

State Model	Organizational Context	Evaluation Model and Purpose
Sovereign state	Parliamentary chain of command	*Peer review model* in order to strengthen professional quality *Client model* in order to strengthen responsiveness *Stakeholder model* in order to increase stakeholder influence
Autonomous state	Professions	*Goal-attainment model* in order to strengthen sovereign state *Client model* in order to strengthen responsiveness *Stakeholder model* in order to increase stakeholder influence
Negotiated state	Interested parties	*Goal-attainment model* in order to strengthen sovereign state *Peer review mode*l in order to strengthen professional quality *Client model* in order to strengthen responsiveness
Responsive state	Clients	*Goal-attainment model* in order to strengthen sovereign state *Peer review model* in order to strengthen professional quality *Stakeholder model* in order to increase stakeholder influence

Recommendations: A Summary

So, several different types of reasoning related to the choice of evaluation models are available. Using different types of reasoning produces different recommendations. Table 6 summarizes the recommendations and the criticisms of the various types of reasoning. The different sets of recommendations highlight different principles for evaluation design, while they are also based on fundamentally different logics.

Recommendations along the lines that design should be determined on the basis of the purpose of the evaluation are founded on a classic goals–means rationale. Recommendations that design should be determined by the characteristics of the evaluated object are either based on a contingency-based goals–means rationale or on a context-based value rationale. Recommendations that design should be determined by the problem that the evaluated object is meant to solve can be characterized as based on a programme theory rationale. The programme theory rationale has a huge potential if it is possible to specify the important variables, but it becomes unrealistic if the programme theory is implicit and the number of variables large.

457

Table 6. Criteria for Choice of Evaluation Model: A Summary

Recommendation	Foundation	Criticism
The *purpose* of evaluation (control vs learning) should determine choice of model	Goals–means rationale	Is it realistic (is the purpose of evaluation clear?)? If the purpose is control: does reliable effects measurement exist?
The *object* of evaluation should determine choice of model:		
a) the possible	a) Contingency-based goals–means rationale	a) Limited design guidance, as most evaluation situations will merely result in comparative evaluation
b) the legitimate and just or a change strategy	b) Context-based value rationale	b) Conservative? Not appropriate recommendation if the evaluation is part of a strategy for reform
The *problem* that the evaluated object is to solve should determine the evaluation design	Programme theory rationale	Is it realistic when interventions are complex and integrated, and the number of variables large?

Reflections on Other Factors Influencing Evaluation Design

It is not known to what extent the theoretical, methodological rational design considerations and reasoning presented affect the practice of evaluation design. As evaluation design is most often a process characterized by task uncertainty, other 'logics of action' than the rational forms discussed above may be expected to affect the design process. There are four potential hypotheses. Evaluation design process may:

- take the form of negotiation;
- reflect logics of appropriateness rather than logics of rationality;
- be influenced by standard operating procedures;
- reflect the 'response repertoire' of evaluation sponsors and evaluators.

Negotiation Processes
A number of actors are often involved in the design process. It may prove helpful to visualize the process as a funnel. Unless evaluation in the area has become a routine, the room for manoeuvring when determining the design will be relatively large early in the phase. But gradually the choice of model, analytical dimensions and assessment criteria narrow. Different actors can have different interests in and requests for the design that are often objects of negotiation.

In light of the state model thinking discussed (Table 4), one can imagine that a concrete design process involves actors that, due to factors such as their position at different levels in the political-administrative system, have different

458

requests for the evaluation design. It is, for example, conceivable that those ordering an evaluation, such as a department or an agency, are affected by the chain of reasoning related to the sovereign state model. Therefore, they might advocate a design based on the goal-attainment model, while those subject to evaluation might be influenced by the reasoning characterizing the autonomous state model and therefore advocate a design influenced by the peer review model. In this situation the design process may transform into a process of negotiation. Different outcomes of such processes are imaginable. The most influential actor(s) may win the struggle over the design process. But it is equally possible that compromises will characterize the process, leading to a spacious design combining multiple models and thus allowing for the interests of all essential actors.

Appropriateness

Another logic of action related to negotiation, but less directly, is the logic that design should be adapted to what the designers expect to be appropriate given the image of the actor and the situation (March and Olsen, 1989). A way of adapting to 'the appropriate' can be to imitate an evaluation design that is considered successful and transfer it between countries, between related policy areas or merely between concrete evaluation assignments within the same policy area (Hansen, 2000). Following this logic of adaptation will reduce the level of conflict. Negotiation will be rendered superfluous in that the commissioners of the evaluation and the evaluators 'voluntarily' and flexibly construct the evaluation design so as to match the wishes, values and interests they expect, observe and experience as related to the assignment.

What is Usually Done

A third logic of action which may influence design is related to 'usually we . . .' statements. Both evaluation sponsors and evaluators fall back on past experiences and allow their choice of design to be influenced by earlier successes or at least the avoidance of serious problems. In this way routines reduce uncertainty in the design phase. Evaluation design develops into a standard operating procedure (SOP).

What Can be Done

A variation on this is the 'what we can do' logic of action. Here uncertainty is reduced by virtue of the fact that the designers simply resort to doing what they are competent in. Uncertainty may not be an issue, the assumption being that management of the assignment is based on the professionalism and evaluation paradigm into which the designers have been socialized. The evaluation designers have, by virtue of their education (among other things), a 'response repertoire' (Weick, 1969). The design process may therefore be characterized by the projection of competence.

Competence projection may occur when the body commissioning the evaluation and/or the evaluators share educational background in a profession of well-developed evaluation traditions. For example, economists would probably

most often adopt economic models, political scientists goal-attainment or effect models, while organization sociologists would probably adopt process and/or actor models.

Conclusion

I have argued that the literature on evaluation design discusses and recommends different logics or kinds of reasoning and different criteria for choosing an evaluation design. Some authors recommend that evaluations be designed in terms of their purpose. Others recommend that evaluations be designed according to what is possible, what is legitimate and just or what change is planned in the characteristics of the evaluated object. Yet others recommend that evaluations be designed on the basis of analyses of the problem that the evaluated object is intended to solve. In other words, recommendations are influenced by a goals–means rationale, by a contingency-based goals–means rationale, by a context-based value rationale or by a programme theory rationale. Overall, the recommendations presented in the literature are ambiguous, and the design-programme theory of evaluation design evidently needs clarification.

This situation in itself must be assumed to contribute to the practice of leaving room for other logics of action to influence the design process. Four possible alternative logics of action have been outlined. The design process may assume the character of a process of negotiations, of accommodation to that which is regarded as appropriate, of establishing routines or of the projection of competence. We lack systematic knowledge about how significant these logics and processes are in practice. Presumably variations occur, for example, between policy areas. We thus need empirical analyses of design processes tied to evaluation.

Note

1. Vedung (1997) employs the term 'side-effects model' for a model which arguably places itself between the goal-attainment model and the effects model. The idea in a side-effects model is to uncover the attainment of goals as well as side effects. This model distinguishes between effects in terms of goals and effects outside of the goal area, the foreseen as well as the unforeseen, negative as well as positive.

References

Birckmayer, Johanna D. and Carol Hirschon Weiss (2000) 'Theory-Based Evaluation in Practice', *Evaluation Review* 24(4): 407–31.

Cameron, Kim S. (1986) 'Effectiveness as Paradox: Consensus and Conflict in Conceptions of Organizational Effectiveness', *Management Science* 32(5): 539–53.

Fetterman, David M. (1996) 'Empowerment Evaluation: An Introduction to Theory and Practice', in David M. Fetterman, Shakeh J. Kaftarian and Abraham Wandersman (eds) *Empowerment Evaluation: Knowledge and Tools for Self-Assessment and Accountability*. Thousand Oaks, CA: SAGE.

Foss, Olaf (2000) 'Evaluering av offentlig virksomhet', in Olaf Foss and Jan Mønnesland (eds) *Evaluering av offentlig virksomhet. Metoder og vurderinger*, pp. 17–51. Oslo: NIBR.

Hansen, Hanne Foss (1989) 'Moderniseringens effektivitet', *Nordisk Administrativt Tidsskrift* 70(2): 189–212.

Hansen, Hanne Foss (1999) 'Organisatorisk effektivitet', in Torben Beck Jørgensen and Preben Melander (eds) *Livet i offentlige organisationer*, pp. 277–313. Copenhagen: Jurist- og Økonomforbundets Forlag.

Hansen, Hanne Foss (2000) 'Kun i mørke er alle katte grå: refleksioner over copycat-praksis og dennes effekter', *Politica* 32(3): 245–61.

Hansen, Hanne Foss (2003) *Evaluering i staten*. Frederiksberg: Samfundslitteratur.

Hansen, Hanne Foss and Finn Borum (1999) 'The Construction and Standardization of Evaluation: The Case of the Danish University Sector', *Evaluation* 5(3): 303–29.

Hansen, Hanne Foss and Birte Holst Jørgensen (1995) *Styring af forskning: kan forskningsindikatorer anvendes?* Copenhagen: Samfundslitteratur.

Jørgensen, Torben Beck (2003) 'Værdier i harmoni, konflikt og forandring', in Torben Beck Jørgensen (ed.) *På sporet af en offentlig identitet*, pp. 63–80. Aarhus: Aarhus Universitetsforlag.

March, James G. and Johan P. Olsen (1989) *Rediscovering Institutions*. New York: Free Press.

Mintzberg, Henry (1983) *Structure in Fives: Designing Effective Organizations*. Englewood Cliffs, NJ: Prentice-Hall.

Owen, John M. and Patricia J. Rogers (1999) *Program Evaluation: Forms and Approaches*. London: SAGE.

Pawson, Ray (2002) 'Evidence-Based Policy: The Promise of "Realist Synthesis"', *Evaluation* 8(3): 340–58.

Pawson, Ray and N. Tilley (1997) *Realistic Evaluation*. London: SAGE.

Premfors, Rune (1989) *Policyanalys*. Lund: Studentlitteratur.

Rossi, Peter H., Mark W. Lipsey and Howard E. Freeman (2004) *Evaluation: A Systematic Approach*. Thousand Oaks, CA: SAGE.

Scott, W. R. (2003) *Organizations: Rational, Natural and Open Systems*. Englewood Cliffs, NJ: Prentice-Hall.

Scriven, Michael (1973) 'Goal-Free Evaluation', in Ernest R. House (ed.) *School Evaluation: The Politics and Process,* pp. 319–28. Berkeley, CA: McCutchan.

Scriven, Michael (1991) *Evaluation Thesaurus*. Newbury Park, CA: SAGE.

Scriven, Michael (2003) 'Evaluation Theory and Metatheory', in Thomas Kellaghan, Daniel L. Stufflebeam and Lori A. Wingate (eds) *International Handbook of Educational Evaluation*. Dordrecht: Kluwer Academic Publishers.

Seglen, Per O. (1992) 'The Skewness of Science', *Journal of the American Society for Information Science* 43(9): 628–38.

Seglen, Per O. (1994) 'Causal Relationship between Article Citedness and Journal Impact', *Journal of the American Society for Information Science* 45(1): 1–11.

Stame, Nicoletta (2004) 'Theory-Based Evaluation and Varieties of Complexity', *Evaluation* 10(1): 58–76.

Stufflebeam, Daniel L. (2000) 'Foundational Models for 21st Century Program Evaluation', in Daniel D. Stufflebeam, George F. Madaus and Thomas Kellaghan (eds) *Evaluation Models: Viewpoints on Educational and Human Services Evaluation*, pp. 33–84. Boston, MA: Kluwer Academic Publishers.

Thompson, J. D. (1967) *Organizations in Action*. New York: McGraw-Hill.

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

Weick, Karl E. (1969) *The Social Psychology of Organizing*. Reading, MA: Addison-Wesley.

461

HANNE FOSS HANSEN, PhD is professor in public organization and administration at the Department of Political Science, University of Copenhagen. She has written a number of books and articles about organization in the public sector, particularly related to issues of effectiveness and evaluation. Please address correspondence to: Department of Political Science, University of Copenhagen, Oester Farimagsgade 5, P.O. Box 2099, DK – 1014 Copenhagen K, Denmark. [email: hfh@ifs.ku.dk]

462