

Chapter 5

There Is More to Methodology than Method

Louise Potvin and Sherri Bisset

There is much more to evaluation than collecting, analyzing, and interpreting scientific data in order to compare the outcomes of various treatments. For the past 40 years, method-related discussions in the field of program evaluation have evolved to include models and reflections on the complex and multiple roles associated with the practice of evaluation. In fact, for Shadish, Cook, & Leviton, (1991), the knowledge basis which pertains to the practice of evaluation must consider issues related to the evaluator's roles as well as to the design of evaluation. Thus, evaluation practice requires both the methodological and technical competencies for systematic inquiry, in addition to a whole set of interpersonal and negotiation skills, identified by Brown (1995) as ranging from pedagogical to political. Conceiving evaluation as a practice, as we do in this collection of essays, is based upon the premise that evaluators are more than good and rigorous scientists, implementing empirical inquiry devices to study programs and interventions. While there is clearly more to the evaluator's role than data-related activities, how do these two aspects of practice which are inherently part of program evaluation, come together and build knowledge of evaluation practice? To address this question, we propose revisiting the teleological, epistemological and ontological foundations upon which evaluation roles are defined.

Consistent with the orientation of this book, we define evaluators' roles vis-à-vis a program as framed by one's evaluation practice. We suggest that evaluation practice does not however simply represent a repertoire of roles from which the evaluator may (more or less) arbitrarily choose from in order to define themselves and their evaluation activities (i.e. their methodological tool kit). In this chapter we argue that the evaluator's role vis-à-vis programs stakeholders and the approach taken to identify, describe, and measure a program and its effects, form an evaluator's practice, which is consistent and coherent across the contexts of their work. We will also argue that a practice, like a paradigm, constitutes an organized rationality (Crozier & Friedberg, 1977) common to groups of individuals that allow

L. Potvin

Department of Social and Preventive Medicine, Université de Montreal, Montreal, QC, Canada

those groups to identify, structure, interpret and solve practical problems identifiable through the lens of a practice or paradigm (Kuhn, 1962).

Four Dimensions of Practice

Methodology is one of four dimensions that characterize a practice (Levy, 1994). Practices like paradigms have a coherent organization such that, as taken together these four dimensions are highly related to each other and constitute a rational and coherent set of propositions that connect a practitioner to the world through her practice. Those four dimensions are labeled: teleological, ontological, epistemological, and methodological.

The teleological dimension relates to the ultimate goal of the practice: what the practitioner is setting herself to achieve through her work and activities. It is this dimension that defines intentionality and relates to the overarching project and vision underlying the practice. It provides meaning to action by identifying the possible worlds the practice is contributing to. This dimension is often taken for granted and rarely openly discussed and critically reviewed in the field of evaluation. Typical of a practical field, however, evaluation results are often described as having to be usable and contributing to some transformation project (Mark & Henry, 2004). Evaluation is often defined as providing scientific data in support of decision-making. Such an instrumental use of evaluation results in program-related decisions constitutes the ultimate criteria by which a great majority of evaluators describe the *raison d'être* of their field (Preskill & Caracelli, 1997). Such an instrumental use however is rarely attained (Patton, 1997) and many theorists of evaluation have proposed other types of uses for evaluation results (see Hartz, Denis, Moreira, & Matida, 2008; Chapter in this book for a more complete discussion of evaluation use) such as enlightenment, that is the contribution of evaluation results to theoretical explanations about the functioning of the world (Weiss, 1998). For Mark, Henry and Julnes (2000) all those specific uses of evaluation results can be subsumed to contribute to social betterment in light that “even in the absence of direct use, evaluation results often appeared to help shape people’s assumptions, beliefs and expectations, and in turn they appear to influence subsequent decisions about programs and policies, sometimes distant in time and place from the original evaluation” (Mark et al., 2000, p. 22). For these authors in order to contribute to social betterment, evaluators take the following responsibilities with regard to their study results. First, evaluators determine which results can best contribute to support the deliberations, decisions and actions carried out by institutions, since institutions represent the legitimate agents of societal regulation and transformation in democratic societies. Second, they take responsibility for the quality of information derived from the actual evaluation studies. Third, they ensure that results are disseminated in the relevant practitioner and decision-maker networks.

The second dimension of practice is related to its object. Considering that a practice is the transformative work of a practitioner upon her world, this work focuses on a specific class of objects. There is no practice without the objects which a

practice seeks to regulate, reproduce or transform. This is the ontological dimension of practice. With the exception of Weiss (1998) there is very little discussion about the ontological reality of the programs and policies that are usually presented as objects of evaluation (Potvin, Gendron, & Bilodeau, 2006). Almost nobody ever asks questions about what programs are made of. As discussed in Chapter 3 of this book (Potvin & McQueen, 2008), most discussions and representations of programs are devoid of actors, taken in the sense of agents who exercise causal power in a situation; actors whose actions induce a reaction in other actors. In the rare occasions where actors are discussed together with material objects, such as in Weiss (1998), the articulation and connections between human and nonhuman program elements are not well developed. Representing programs as series of actions or events in the absence of actors who operate those actions puts the onus of the action in the technical and material dimensions of programs. In this chapter we argue that evaluators' conceptions about the nature of programs and what constitutes their reality bear fundamental influence on evaluation practice, and ultimately define the other dimensions of practice.

Defining practice as the work of a practitioner on an object implies the presence of a rapport between the practitioner and the class of objects that characterizes the practice. The epistemological dimension of practice is about this relationship between practitioners and the object of her practice. It asks the question of the kind of rapport that should exist between a practitioner and the world in order for the former to influence transformations or regulation on the latter. For an evaluator this dimension is about the relationship between herself and programs or policies to evaluate. Often simplified as an opposition between outside or inside evaluation (e.g. participatory versus non-participatory research, participant versus non-participant observation), a proper answer to this question requires the evaluator as a subject to position herself with regard to the program. Evaluation practice can range from a simple subject-object rapport in which the evaluator as a subject controls the circumstances of the program ideally conceived as devoid of power in the relationship (evaluator's decisions are not influenced by the program), to a much more complex subject-subject-project rapport in which both the evaluator and the program are active agents in the evaluation, and in which, this rapport is also conceived as being an active ingredient in the evaluator's practice. We have shown that when evaluation is designed to be responsive to the various phases through which a program evolves, both the program and the evaluation influence each other and become increasingly ingrained in the context in which the program was initially developed (Potvin, Cargo, McComber, Delormier, & Macaulay, 2003). For the research base practice of evaluation, this epistemological dimension which ultimately poses the question of the status of the evaluator with regard to the program, is concerned with how knowledge about programs is possible given their nature.

Finally the methodological dimension of a practice refers directly to the type of actions practitioners undertake to achieve their goal. For an evaluator, the methodological dimension of practice is primarily, but not exclusively, related to how knowledge about programs is produced but it cannot be reduced to this sole aspect. Indeed, to the extent that the finality pursued by an evaluation expands beyond

the mere production of scientific knowledge, the methodological dimension of evaluation practice may include a wide range of actions that pertain principally to evaluability assessment which defines the evaluation question and project (Thurston, Graham, & Hatfield, 2003; Thurston & Potvin, 2003) and to knowledge use (Mark & Henry, 2004; Patton, 1997).

Three Ideal-Types of Evaluation Practice

A practice is the operationalization of a coherent set of positions regarding the four dimensions described above. This coherence is the foundation of practice, and is observed as a consistency and continuity of practice across time and space. Although it is conceivable that the various combinations of answers practitioners provide to those four questions could produce an infinite variety of practices, we believe that there exist inherent correlations between these dimensions, thus limiting variations in the types of evaluation practice. In this respect, Weber's notion of ideal-type is useful (Weber, 1952). An ideal-type is a mental construction of a social phenomenon that combines and simplifies characteristics usually not to be found in any single instantiation, but which provides an exemplary case representative of a whole category of social phenomena. There exist, many taxonomies of evaluation ideal-types, depending on the features one wants to highlight. In this section we present three ideal-types of evaluation practices that emphasize the methodological consequences of three coherent evaluation paradigms as presented in Table 5.1.

Note that some well-known evaluation ideal-types practices such as utilization-focused evaluation (Patton, 1997), or participatory evaluation (Springett, 2001) are

Table 5.1 Three ideal-types of evaluation practice

	Teleological	Ontological	Epistemological	Methodological
Evaluation as Experimentation	Testing causal hypotheses about programs' effects on context	Programs as technical entities with an objective reality	Objective distance between evaluator and programs	Experimentation; ruling out plausible rival hypotheses
Evaluation as Negotiation	Improve program by increasing programs actors awareness about their actions	Programs as representations that describe and/or guide actors' actions	constructivist subject–subject co-construction of program representation	Constructing program's theory through actors' representations and hermeneutic circles
Evaluation as Organized Reflexivity	Understanding the mediating role of programs in social transformations	Programs as systems of actions that connects entities	Reflexive transformation of subject-object-project	Following programs actions and the connections they operate

not directly circumscribed within our typology, mainly because such practices are not primarily discussed in terms of their methodological implications. The former is generally presented with regard to the teleological dimension of evaluation, i.e., its use by decision-makers, whereas the latter is mostly presented in relation to the epistemological question of the relative value of various forms of knowledge. In this light however, participatory evaluation can also be understood as a form of negotiation.

The main caveat about the use of ideal-types to characterize health promotion evaluation practice is that because they do not necessarily exist in reality, ideal-types are simplistic representations. As a heuristic device, an ideal-type cannot render all the subtleties and nuances of the social reality it represents. It is called upon mainly to contrast social phenomena by exacerbating their differences on a limited number of dimensions. A practice is much richer than any ideal-type that can be constructed to represent it. This is why this chapter has to be complemented by the chapters composing the fourth section of this book, in which colleagues from the Americas, North and South report and discuss practical issues of conducting health promotion evaluations that also contribute to the health promotion agenda.

Evaluation as Experimentation

According to Pawson and Tilley (1997) this ideal-type of evaluation practice has been the hallmark of evaluation since the early time of this field. "Underlying everything in the early days was the logic of experimentation (. . .) The practitioner, policy adviser, and social scientist are at one in appreciating the beauty of the design. At one level it has deepest roots in philosophical discourse on the nature of explanation as in John Stuart Mill's *A system of Logic*; at another it is the hallmark of common sense, ingrained into advertising campaigns telling Washo is superior to Sud" (Pawson & Tilley, 1997, p. 4).

Evaluation as experimentation practice is characterized by the finality of testing causal hypotheses about program's effects in the environment into which they are implemented. This finality is deeply rooted in the field of evaluation. As early as 1967, in one of the first books published with the word evaluation in the title, Suchman (1967) explained that evaluation is ultimately a hypothesis testing mechanism whereby demonstrating a positive effect of a program designed to solve a problem is providing evidence of the validity of the theory underlying the program. In practice, as hypothesis testers, evaluators are also often academic researchers, deeply involved in program design and planning. Programs as treatments are seen as devices to test theoretical propositions about disease etiology. This was the case with the first cardiovascular disease community intervention trials such as the Stanford Five City Project (Farquhar et al., 1985), the Pawtucket Heart Health Program (Carleton, Lasater, Assaf, Lefebvre, & McKinlay, 1987), and the Minnesota Heart Health Program (Blackburn et al., 1984). In these programs, as in those that were designed and evaluated in their aftermath according to the same evaluation

paradigm, there is a strong involvement of researchers/evaluators in the definition and design of the program that usually stops just short of implementation, keeping evaluation at arms length from implementation issues, but not from the stakes of demonstrating a program effect.

As exemplified by those programs designed by academic researchers and which encompass the most update scientific knowledge about disease etiology and prevention, in evaluation as experimentation, programs are conceived of as generalizable solutions to objectively defined problems. The origin of the program lies outside the problematic situation, that is, the program is assumed to come after the problem and after potential solutions have been named. It is this externality that bears generalizability. Programs are thus essentially conceived of as technical solutions that incorporate scientific knowledge in response to a problematic situation, without reference to the social actors involved with the problem, its definition, or its solution. The paradigm of reference is that of drug development (Flay, 1986). In this paradigm programs are made of material objects arranged according to a stable and predictable set of procedures. Contextual and implementation variations are noise to be eliminated or controlled for as much as possible (see Chapter 17 of this book, Poland, Frohlich, & Cargo, 2008) for an in-depth refutation of this proposition). Here, the causal mechanisms that produce changes are not understood as being between the actors who operate the program, or in the relationships that they develop with other actors and non-human components of the implementation context. Instead, actors are instruments whose role is defined in the program's description and logic model.

When programs are primarily conceived as technical arrangements, usually, the agency of the concerned actors is likewise regarded as being quite limited as their interactions with the technical program entities are generally pre-defined by a set of rigid program instructions and rules. This is to ensure maximum fidelity in program implementation and theoretically maximum program effects. Indeed, in the experimentalist language, people who are to benefit from programs effects are often described as program targets. This ballistic metaphor is not benign and this on two counts. First targets are situated exclusively at the receiving end of a transaction. They are meant to be hit by something (i.e., the magic bullet). Second, and most importantly, although targets may be moving, it is not what makes them move that constitutes relevant information. The only important fact about a target's movements is its relative position (distance and direction) in relation to that of the person aiming at that target.

In evaluation as experimentation, relationships between the evaluator and programs, including staff and targets, are usually conceived as being one way. The more specific the hypothesis to be tested by the program and its evaluation, the less flexibility there is in the expression of agency by actors involved. This objectification of programs and their components ideally maximizes control over the treatment and potential confounders. As knowing subject, the evaluator's actions are minimally influenced by the object (program) under study. Conversely, one key hypothesis for external validity – the capacity to generalize the results to other settings and targets –

is that the program is not responsive to evaluator's actions. In other words an evaluated program is not different from a non-evaluated program. The epistemological dimension is thus characterized by maximizing distance and objectivity between the evaluator and the program evaluated.

In this ideal-type practice, methodological choices are fundamentally governed by the experimentalist paradigm. Knowledge is enhanced by one's capacity to manipulate the conditions of production of a phenomenon and reproduce this phenomenon by manipulating the conditions of its emergence, reducing these conditions to the bear essentials. The essence of laboratory science is to isolate, as perfectly as possible, a putative cause and its effect from all other confounding sources. This is done by closing and isolating the experimental situation. Cook & Campbell (1979) have discussed at length the conundrum associated with the emulation of laboratory conditions in research conducted in real life settings where this closure is hardly feasible. In their view, the work of the evaluator in such situations is to try to impose closure to the situation while maintaining the integrity of the experimental treatment. The randomized control trial is one device that allows closure on several but not all aspects of the experimental situation in real-life condition. In fact, according to Campbell (1984), the hard work of an evaluator can be summarized as trying to rule out plausible rival hypotheses in order to remain only with the program under study as an explanation of the observed difference between those who were exposed to it and those who were not. This can be done either by designing studies that automatically rule out some known threats to internal validity (Campbell & Stanley, 1963), like using a control group in order to rule out the possibility that the observed pre-post difference in the exposed group could be due to maturation effect. This can also be done by a post hoc documentation of the low plausibility of such hypothesis, or by statistical control (Cook & Campbell, 1979).

The evaluation as experimentation practice is characterized by a strong concern with the conception and implementation of research devices that ensure optimal internal and external validity of the evaluation results. Ideally, this includes maintaining a proper distance with the program and ensuring that the work and activities serving the evaluation do not interfere with the program's integrity.

Evaluation as Negotiation

Contrary to the previously discussed evaluation as experimentation where emphasis is placed upon the technical identity of the program, the emphasis in evaluation as negotiation is on the social actors who are brought together and form a programmed social space. Recognizing that programs are essentially social systems in which actors' actions and interactions remain the main dimension of interest, evaluators as negotiators focus their practice on the people who are directly concerned by the program. Methodological preoccupations are mainly related to compiling program representations from the perspectives of the various groups of actors. This primacy

of social actors and social processes opens up a royal path to social constructivism as a paradigm for program evaluation.

This evaluation ideal-type practice has been mainly, but far from exclusively, advocated by Guba & Lincoln (1989) as the “Fourth Generation Evaluation”. Indeed, although the constructivist perspective of Guba and Lincoln is probably the one most acknowledged and cited in the health promotion evaluation literature, some of the well known pioneers in the field of evaluation have also founded their approach on constructivist epistemology. The work of evaluators such as Wholey & Newcomer (1989) or Stake (1975) are well known in the field of evaluation. Both are associated with the fundamental idea that programs as implemented usually differ from the original program planners’ intentions; programs are constructed as they go through the various interactions among and between groups of associated actors. One of evaluation’s key role is thus to highlight this process by supporting program improvement or simply defining the program on its own terms. By emphasizing the necessity for evaluators to get closer to the action wherever it occurs, both contributed (with others) in the early 1970s to the decline of the quasi-experimental paradigm monopoly over the field of evaluation. Building on the notion that programs are social constructions, their most important contribution was to successfully argue for pluralism of methods and relativism of values when assessing programs.

While there are numerous blends of social constructivism the one that is most often found to underlie constructivist evaluations is rooted in the work, and scholarly tradition of the Chicago School of Sociology, mainly through the work of Anselm Strauss and his methodological propositions, which are embedded within grounded theory type inquiry (Glaser & Strauss, 1967). Although the Chicago School of Sociology is associated with a great number of theoreticians and methodologists in the field of sociology, one of its most critical and relevant innovation for evaluation is the so-called symbolic interactionism. The first and most fundamental proposition of symbolic interactionism is that humans act essentially as a result of their perceptions and interpretations of the meaning they ascribe to others’ actions. The second and corollary proposition is that it is through interactions with others that these meanings develop, are transformed, and can be manipulated (Le Breton, 2004). The main driver of action thus is the representation of social situations and those representations are shaped through the social processes of interacting within those situations.

In developing their approach to the “Fourth generation evaluation” Guba & Lincoln (1989) borrowed extensively from symbolic interactionism, from the methodology of the grounded theory, and from the work of Robert Stake (Abma, 2005). They argue that programs can only be understood through the representations that various actors develop and act upon through their social interactions with other program actors.

In both Guba and Lincoln’s and in Stake’s conception, the evaluators’ main task is to support the development of a shared program representation among actors and this is essentially done through negotiation. The teleological dimension in evaluation as negotiation practice is thus to contribute to program improvement as it is being implemented, acknowledging that discrepancies between the program as

it is planned, and the program as it is practiced are unavoidable. Here, programs are expected to improve with an elaboration of a common program representation among program's stakeholders, whereby, a common representation is presumed to improve the program through a better coordination and alignment of program's stakeholders' actions. It is the evaluators' main task to develop this shared representation, primarily by feeding back their assessments to program actors thereby permitting the latter to revise their representations in line with those of other program actors.

In line with symbolic interactionism, the ontological dimension of this ideal-type practice lies in actors representations of the program. Thus programs do not have an objective reality in the sense of the primacy of material and technical objects. It is the representations program actors construct of those objects and the symbolic meaning they ascribe to them through their interactions that are orienting individuals' actions in programs' social spaces. So in any program, theoretically, there exist as many representations as there are individuals involved. The evaluator's role is to attempt to reveal those representations and to help achieve consensus among program actors.

This task requires that the evaluator interacts closely with program actors. The epistemology is constructivist, meaning that both the evaluator and program actors are subjects in this construction. There is no prescribed distance between the program and the evaluator. Indeed the program's shared representation resulting from the evaluation can only be collectively constructed and the evaluator's central role is to design and implement the mechanisms through which actors' representations can directly or indirectly confront with one another. For example, mechanisms in the form of feedback derived from discourse, documentation or observational analyses, provide a means for program actors to situate their own interests and interpretations of program events in relation to those of others.

Although there exist variations in the specific technical devices used to collect and analyze data, the methodological dimension in evaluation as negotiation ideal-type practice are essentially concerned with producing the most credible and trustworthy program representation or program theory. In many cases evaluators have referred to the grounded theory methodology as developed by Glaser and Strauss (1967) and later by Strauss and Corbin (1998) in order to inductively create a shared representation from the specific individuals' program representation. Specific data collection and analysis techniques insure a continuous confrontation process between the theory that is emerging from the data and the empirical phenomena which are illuminated by the theory. Guba and Lincoln's (1989), main practical and methodological addition lies in the proposition that Hermeneutic Dialogic Circles should be implemented in order to construct this shared representation. As ultimate negotiators and consensus brokers, evaluators orchestrate dialogs between program actors in order to create a joint construction that comprised as many elements as possible from each individual's program representation.

In health promotion several variations of this ideal-type of evaluation practice have been described as having utmost relevance given the nature of health promotion. Presenting Stake's theory of responsive evaluation, Abma (2005) argues that "Responsive evaluation is not only responsive to the unique feature and emerging

ideas in the field of health promotion, it is also synergistic with health promotion” (Abma, 2005, p. 287).

In particular, this approach is touted as having a potential to redress the social processes which maintain health inequalities through the creation of equal partnerships with full participation among health or research experts and community or lay actors. Evaluations then ideally showcase the working relationships or partnerships between various actors as the object of study. Evaluations of community-based programs, for example, describe a set of actors with a range of pre-occupations with community, non-governmental, governmental or research needs as they come together to define a problematic situation and devise corrective strategies. Here, the role of the evaluator is to attempt to trace this process by collecting and interpreting qualitative data in order to capture the perspectives, opinions and lived experiences among the various actors. In the published literature, analytic results from such studies present representations in aggregate where the various actor groups are identified at some level of collective. A collective may be quite general, whereby one overall or encompassing representation is attributed to all the program actors (Lantz, Viruell-Fuentes, Israel, Softley, & Guzman, 2001). Alternatively, representations may be distinct or specific to the various actor groups according to a pre-defined group membership (e.g., community member, health professional) (Schulz et al., 2001). In addition to being organized according to the identity of program actors, data may be interpreted in relation to its correspondence with a stage in the program planning and evaluation (Farquhar, Parker, Schulz, & Israel, 2006) or to a set of concepts which are associated with the guiding principles of the program (e.g., participatory research) (Savage et al., 2006).

More specifically, one initiative, the Detroit Community-Academic Urban Research Center (URC) provides an example where the evaluation plays a role of negotiator by building a collective representations of the partnership building process between a range of actors. Evaluations of this initiative were described at the regional level, where community and academic actors formed a committee (i.e. URC), applied for, and dispersed, funds to local initiatives, and also at the local level where initiatives or projects addressed a particular problem within a particular place. Evaluations at both the regional and the local levels were based upon the representations of the actors with respect to the partnership building process, as well as, the operation of services or acquisition of resources.

The East Side Village Health Worker Partnership (ESVHWP) was one of twelve local projects which was funded through the Detroit Community-Academic URC between 1995 and 1999 (Lantz et al., 2001). Evaluation of this project relied upon the derivation of program representations based upon informant interviews, observations and document review in relation to four stages of program planning process (Farquhar et al., 2006). While data was collected over time, analysis of a given data set within one time period was interpreted according to its reference to four stages of planning (i.e., assessment and problem analysis, goals and objectives, design and implementation, evaluation). The evaluation furthered its role as a negotiation process by feeding back constructed representations to the various program actors, which served both as part of the learning and informed decision-making process

and also as a source of legitimization for the lived experiences and representations among the interventionists (i.e., health workers) and steering committee member. For example, by feeding back interpretations of the health worker meetings, the evaluation validated the presence of four distinct problems upon which health workers divided themselves into four subgroups, each with its own focus and set of activities. This approach is somewhat distinct to hermeneutic circle dialogue in that it did not impose a consensual problem statement.

Other evaluations of the Detroit Community-Academic URC (Lantz et al., 2001) and also the ESVHWP (Schulz et al., 2001) identified a diverse range of actors who came together to form a research centre (i.e., URC Board) or a steering committee. Evaluation of the URC Board was based upon a collective representation of the internal and external strength of the partnership, namely how well the partnership was able to move forward as a group toward goals, and how the group encountered conditions which either facilitated or acted as barriers to their public health research goals. Evaluation results were organized into categories which identified the overall satisfaction, perceived benefits, facilitating factors and barriers toward goal attainment. For example, issues which were identified as concerns to the group were time management, resource distribution and balancing community and research needs. Examples of facilitating factors toward goal attainment included community representation on URC Board, trust, relevant knowledge, and organizational support. Alternatively, evaluations derived specifically from the informant interviews provided by the Village Health Workers and the steering committee as part of the ESVHWP, aimed at identifying how the partnership contributed to improving the research, program activities, community relations and participation from community, academic and practice organizations or institutions. Interviews were interpreted according to these over-riding themes and subcategories specific to the themes were induced from the interviews and results were presented as an overall collective representation across both the steering committee and the Village Health Workers. For example, improvements in community relationships were described by stronger social networks among Village Health Workers, stronger relationships between them and steering committee members, and stronger relationships among academic, practice, and community-based organizations. Within these subcategories particular issues, such as trust, power differentials and communication, along with some strategies used to address challenges (e.g., annual picnics, retreats) were presented.

Two criticisms are usually associated with this constructivist paradigm of evaluation. The first is the essentially descriptive nature of the results of such an evaluation practice. That is, while it is possible that consensus can be achieved, as conditions change over time and space, so too may a program's ability to achieve consensus among all program actors. As illustrated above, achieving consensus may not be desirable. This description does not therefore permit knowledge to be garnished from the process by which negotiations or compromises between program actors may have taken place in order to arrive at common or agreed upon goal(s). Indeed, in the renewed version of evaluability assessment, the construction of a shared understanding or representation of the program and its components is only the starting

point from which designing the evaluation project begins (Thurston et al., 2003; Thurston & Potvin, 2003). The other criticism relates to the somewhat underestimation of the political influence and power held by evaluators in such processes. As the master negotiator it could be mistakenly understood that the evaluator is neutral in this construction process, which is certainly not the case. The use of formal mechanisms that ensure a fair distribution of roles, resources, and responsibilities in conducting evaluation projects is increasingly seen as a necessary safeguard.

Evaluation as Organized Reflexivity

There is another ideal-type of evaluation practice that is emerging in the evaluation literature and that is slowly making its way into the realm of public health and health promotion program evaluation. Pawson and Tilley (1997) have developed the foundations of a critical realist approach to evaluation (see chapters by Potvin and McQueen (2008); Mercille (2008), and Poland et al. (2008) in this book), based upon the work of the philosopher Roy Bhaskar (1978). “Realist evaluation, as its core, focuses on developing explanations of the consequences of social actions that contribute to a better understanding of why, where, and for whom programs work or fail to work. To this end, realist evaluators place a great deal of emphasis on (a) identifying the mechanisms that produce observable program effects and (b) testing these mechanisms and the other contextual variables . . . that may have impacts on the effects that are observed” (Pawson & Tilley, 2004, p. 359). Working under the realist assumption that programs are but one among many systems of actions that operate simultaneously at any given moment in a context, a realist evaluator’s task is ideally to identify the operative program mechanisms and how they interact with contextual conditions (including other programs) to be associated with specific intended and unintended outcome patterns (Pawson & Tilley, 2004). In a complex world of constant interactions, it is through the connections programs are creating and maintaining with contextual conditions that their impacts can be understood and identified. Thus, program evaluators are essentially designers and operators of organized reflexivity devices through which programs’ actors can understand how programs contribute to transform peoples’ life and conditions.

We propose that the teleological dimension of an evaluation as organized reflexivity practice, is one of understanding the mediations operated by a program in the transformation of the world. The recognition of the inherent complexity of programs and of their operating mechanisms in open systems leads to the critical realist proposition that causality cannot be established solely by the empirical observation of constant conjunctions of events (Bhaskar, 1978). Indeed, in an open system where a multitude of mechanisms are constantly interacting and transforming the system, exact recurring patterns cannot be observed. It is only in controlled conditions where some closure or quasi-closure is exerted on the system that law-like observations can occur. In all other situations, the main role of the scientist is to identify and factor out the effect of irrelevant mechanisms (Cook & Campbell, 1979). In the absence

of such closure, empirical observations can only make sense when related to a theory about mechanism-context-outcome pattern configurations that they contribute to strengthen. As operating mechanisms in a context, a program interacting with other existing mechanisms becomes a mediator in the constant transformation of the context. Its presence changes the normal course of events either by reinforcing, or by tempering with, the impact of other operating forces in place.

In this ideal-type practice, programs are conceived of as systems of actions in which human and non-human actors operate the putative program mechanisms through their actions (Potvin, Gendron, & Bilodeau, 2006). Program's ontological reality is thus one of action and interactions. It is what is done by the various program actors, both human and non-human that contribute to transforming the context in the direction expected by program initiators. Thus, the most relevant program reality rests in the actions undertaken in the program space. These, like all actions, involve two types of interacting realities. One type is tangible and objective, and it consists in the peoples, objects, resources, and physical spaces that situate the action and make it possible. It is usually what is subsumed under the work plan and logic models associated with program development. The second type is symbolic and includes the meanings actors assign to actions and the role those actions implicitly and tacitly assign to actors in relation to action. This is what is usually captured by program representations. A realist ontology of program has to encompass those two types of reality and their interactions.

In order to understand how a program mediates transformations unfolding in a particular space and time, the evaluator has no choice but to follow the actions that are developing and taking place in the name of the program as well as their ramifications and ripples in the program's social context. Following the action is the single most important methodological dimension of a reflexive device. It is through their actions and practice that actors enact their intentions much more than through the discourse they have about their intentions (representations). This is one of the well recognized limits of phenomenology; that one's representation of one's actions is usually blind to the structural conditions that constrain and make this action possible (Bourdieu, 1972). Ideally, it is through description and analysis of the actions that instantiate the program that program actors can critically reflect on the program and rationally make the constant corrections that are necessary to strengthen mechanism-context-outcomes pattern configuration. For example, as practitioners or researchers increase their level of involvement with a group of concerned actors, their representation of what a problem is and how it may be resolved is likely to change. Sometimes, due to the nature of the research, the community or the individuals involved, program plans adapt (Bisset, Cargo, Delormier, Macauley, & Potvin, 2004), other times, however, despite recognizing a need to change or expand upon plans, practitioners latitude is restricted due to an imposed need to stay on course (Hawe & Riley, 2005). Evaluation as organized reflexivity follows why and how transformations of this sort occur, thus providing intelligibility to local adaptations.

The underlying principle upon which a following the action methodology is based, holds that a program is most accurately and comprehensively understood

through an in-depth study of the program performers in action. By following the action of program performers, we can obtain a view of the program from its core functioning. This approach contrasts building an understanding of the program based upon accounts of what the program may aim to accomplish and examples of exemplary activities. In the same way that the sociology of association dismisses ostensive definitions of what is social (Latour, 2006), it can similarly be argued that in defining program upon accounts and examples, we side-step the real work of learning about the upkeep of a program, of what brings the program together and makes it operative.

Two distinctive features can be identified with a methodology based upon following the action. First, it does not limit itself to inquiring about actions which are believed to be part of a causal chain of events leading to a pre-defined set of outcomes. While certainly important, these actions, however, capture a limited portion of what practitioners do and assume they share and are driven by a singular model of achievement. Moreover, actions can only be planned to a certain point, after which they become diverted or develop spontaneously given the conditions which arise.

This last point raises a second unique feature of evaluations which follow the action. Following the action requires that the lived experiences occurring inside and outside the typical boundaries of the program be considered. First, actions taking place outside the physical space and time of program delivery form opinions and impressions which impact upon how program participants and practitioners interact with the program. Second, during the given time space of a program, its form is not solely determined by the instructions and materials which are delivered, but rather by the ways in which the various actors that are present, interact with instructions and materials, and with one another.

Finally, the epistemological dimension of this ideal-type practice emphasizes the reflexive position of the actors involved in the program as well as those involved in the evaluation. Indeed, in following the action as it develops and unfolds, the evaluator is necessarily a relevant part of the program's context. As such, one cannot rule out the possibility that the connections between the program and its evaluation that are necessary in order for the latter to follow the actions taking place within the former, are themselves mediating transformations in the program, the context and also in the evaluation. Not only does an evaluated program differ from what it could have been had it not been evaluated, but evaluation projects also differ according to the programs or versions of programs they are coupled with. So an accurate assessment of the mediating role of the program in contextual transformations, takes into account the transformative role of evaluation on the program. Ultimately, evaluation is an intervention on programs (Mark et al., 2000).

Conclusion

In this chapter we have explored three ideal-type practices for evaluation. We based our exploration of those ideal-types on a paradigmatic conception of practice that supposes four underlying coherent dimensions: teleological, ontological,

epistemological and methodological, as indicated in Figure 5.1 adapted from Gendron (2001). We deliberately chose the methodological dimension as the main point of entry for our exploration. Certainly, entering this exploration through any other paradigmatic dimension would have led to defining different ideal-types. For example, efforts to highlight the teleological dimension of practices would have led to distinguishing utilization-focused (Patton, 1997) from hypothesis testing/knowledge production practices (Cook & Campbell, 1979), and an emphasize on the epistemological dimension would have contrasted participatory from expert-driven practice. Although there should be some correlations between the outcomes of the various points of entry for exploring evaluation practices (it is more difficult to couple an evaluation as experimentation practice with a participatory practice), we do not think that there is an exact correspondence.

By characterizing evaluation practices through their methodological implications we wanted to emphasize the richness and diversity of the various points of entry that can be used for such an exploration, proposing other dimensions than those usually used from which to define practical issues in evaluation. It is important also to note that we did not discuss the qualitative/quantitative dichotomy. For us, this is not a methodological issue but a technical one that pertains to the nature of the data available rather than to the manner by which a subject can derive knowledge about a given object.

We have shown that all of the three ideal-types described in our paper can be found in the health promotion evaluation literature. All have been used with various frequencies and various levels of success, and we think that there are situations in which health promotion is better served by all three of these ideal-types. We want, however, to point at the interest to consider evaluation as organized reflexivity practice for health promotion evaluation. As a field of practice, health promotion is still very young and cannot rely on an important body of solidly established context-mechanism-outcome pattern configurations. There is therefore a great deal of work to be undertaken in order to create a repertoire of interventions that can be operated efficiently in various contexts. Learning as we go and organizing that knowledge

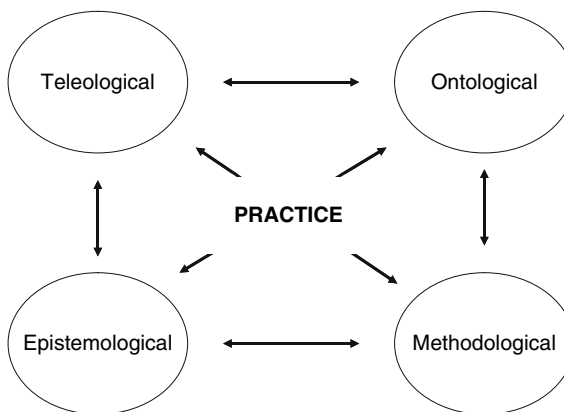


Fig. 5.1 The four inter-related dimensions of practice (Adapted from Gendron, 2001, p. 34)

into coherent intervention theories is certainly a strategy that should be explored and expanded into a proper knowledge base for health promotion. Organized reflexivity as a methodological practice accommodates this strategy of knowledge development. In addition, as shown through several chapters in this book, this ideal-type practice is compatible with many practical and theoretical solutions this book's contributors examined to align the values of health promotion with the scientific rigor of evaluation.

References

- Abma, T. A. (2005). Responsive evaluation: Its meaning and special contribution to health promotion. *Evaluation and Program Planning*, 28, 279–289.
- Bhaskar, R. (1978). *A realist theory of science*. Hassocks, UK: Harvester Press.
- Bisset, S., Cargo, M., Delormier, T., Macauley, A., & Potvin, L. (2004). Legitimizing diabetes as a community health issue: a case analysis of an Aboriginal community in Canada. *Health Promotion International*, 19, 317–326.
- Blackburn, H., Luepker, R. V., Kline, F. G., Bracht, N., Carlaw, R., Jacobs, D., et al. (1984). The Minnesota Heart Health Program: A research and demonstration project in cardiovascular disease prevention. In J. D. Matarazzo, N. E. Miller, S. M. Weiss, J. A. Herd & S. M. Weiss (Eds.), *Behavioral health: A handbook of health enhancement and disease prevention* (pp. 1171–1178). Silver Spring, MD: John Wiley & Sons.
- Bourdieu, P. (1972). *Esquisse d'une théorie de la pratique*. Genève: Ed. Droz.
- Brown, P. (1995). The role of the evaluator in comprehensive community initiatives. In J. P. Connell, A. C. Kubisch, L. B. Schorr, & C. H. Weiss (Eds.), *New approaches to evaluating community initiatives. Concepts, methods and contexts* (pp. 201–225). New York: Aspen Institute.
- Campbell, D. T. (1984). Can we be scientific in applied social science? In: R. F. Connor, D. G. Altman, & C. Jackson (Eds.), *Evaluation studies review annual*, Vol 9 (pp. 26–48). Beverly Hills CA: Sage.
- Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally.
- Carleton, R. A., Lasater, T. M., Assaf, A., Lefebvre, R. C., & McKinlay, S. M. (1987). The Pawtucket Heart Health Program: 1. An experiment in population-based disease prevention. *The Rhode Island Medical Journal*, 70, 533–538.
- Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation; Design and analysis issues for field settings*. Boston: Houghton Mifflin.
- Crozier, M., & Friedberg, E. (1977). *L'acteur et le système*. Paris: Seuil.
- Farquhar, J. W., Fortmann, S. P., Maccoby, N., Haskell, W. L., Williams, P. T., Flora, J. A., et al. (1985). The stanford five city project: Design and methods. *American Journal of Epidemiology*, 122, 323–334.
- Farquhar, S. A., Parker, E. A., Schulz, A. J., & Israel, B. A. (2006). Application of qualitative methods in program planning for health promotion interventions. *Health Promotion Practice*, 7, 234–242.
- Flay, B. R. (1986). Efficacy and effectiveness trials (and other phases of research) in the development of -promotion programs. *Preventive Medicine*, 15, 451–474
- Gendron, S. (2001). *La pratique participative en santé publique, l'émergence d'un paradigme*. Montréal: Thèse de doctorat présentée à l'Université de Montréal. Retrieved in January 2008 from : <http://proquest.umi.com/pqdlink?Ver=1&Exp=01-15-2013&FMT=7&DID=726370931&RQT=309&attempt=1>

- Glaser, B., & Strauss, A. (1967). *The discoveries of grounded theory: Strategies for qualitative research*. Chicago: Aldine.
- Guba, Y., & Lincoln, E. (1989). *Fourth generation evaluation*. Thousand Oaks CA: Sage.
- Hartz, Z. A., Denis, J-L., Moreira, E., & Matida, A. (2008). From knowledge to action: challenges and opportunities for increasing the use of evaluation in health promotion policies and practices. In L. Potvin & D. V. McQueen (Eds.), *Health promotion evaluation practices in the Americas: Values and research*. New York: Springer.
- Hawe, P., & Riley, T. (2005). Ecological theory in practice: Illustrations from a community-based intervention to promote the health of recent mothers. *Prevention Science*, 6, 227–236.
- Kuhn, T. S. (1962). *The structure of scientific revolutions*. Chicago: University of Chicago Press.
- Lantz, P. M., Viruell-Fuentes, E., Israel, B. A., Softley, D., Guzman, R. (2001). Can communities and academia work together on public health research? Evaluation results from a community-based participatory research partnership in Detroit. *Journal of Urban Health*, 78, 495–507.
- Latour, B. (2006). *Changer de société – Refaire de la sociologie*. Paris: La découverte.
- Le Breton, D. (2004). *L'interactionisme symbolique*. Paris : Presses universitaires de France.
- Levy, R. (1994). Croyance et doute: une vision paradigmatique des méthodes qualitatives. *Ruptures, revue transdisciplinaire en santé*, 1, 92–100.
- Mark, M. M., & Henry, G. T. (2004). The mechanisms and outcomes of evaluation influence. *Evaluation*, 10, 35–57.
- Mark, M. M., Henry, G. T., & Julnes, G. (2000). *Evaluation. An integrated framework for understanding, guiding, and improving policies and programs*. San Francisco CA: Jossey-Bass.
- Mercille, G. (2008). A realist approach to synthesizing evaluation results. In L. Potvin & D. V. McQueen (eds.), *Health promotion evaluation practices in the Americas: Values and research*. New York: Springer.
- Patton, M. Q. (1997). *Utilization-focused evaluation. The new century text*. Thousand Oaks: Sage.
- Pawson, R., & Tilley, N. (1997). *Realistic evaluation*. London UK: Sage.
- Pawson, R., & Tilley, N. (2004). Realistic evaluation. In S. Mathison (Ed.), *Encyclopedia of evaluation* (pp. 362–367). Newbury Park, CA: Sage.
- Poland, B., Frohlich, K. L., & Cargo, M. (2008). Context as a fundamental dimension of health-promotion program evaluation. In L. Potvin & D. V. McQueen (Eds.), *Health promotion evaluation practices in the Americas: Values and research*. New York: Springer.
- Potvin, L., Cargo, M., McComber, A., Delormier, T., Macaulay, A. C. (2003) Implementing Participatory Intervention and Research in Communities: Lessons from the Kahnawake Schools Diabetes Prevention Project. *Social Science & Medicine*, 56, 1295–1305.
- Potvin, L., Gendron, S., & Bilodeau, A. (2006). Très posturas ontológicas concernentes à natureza dos programas de saúde: implicações para a avaliação. In: M. L. M. Bosi & F. J. Mercado (Eds.), *Avaliação qualitativa de programas de saúde. Enfoques emergentes*. Petropolis, Brazil: Vozes Editorial.
- Potvin, L., & McQueen, D. V. (2008). Practical dilemmas for health-promotion evaluation. In L. Potvin & D. V. McQueen (Eds.), *Health promotion evaluation practices in the Americas: Values and research*. New York: Springer.
- Preskill, H., & Caracelli, V. J. (1997). Current and developing conceptions of use: Evaluation use topical interest group survey results. *Evaluation Practice*, 18, 209–225.
- Savage, C. L., Xu, Y., Lee, R., Rose, B. L., Kappesser, M., Anthony, S. P. (2006). A case study in the use of community-based participatory research in public health nursing. *Public Health Nursing*, 23, 472–487.
- Schulz, A. J., Isreal, B. A., Parker, E. A., Lockett, M., Hill, Y., & Wills, R. (2001). The east side village health worker partnership: Integrating research with action to reduce health disparities. *Public Health Reports*, 116, 548–557.
- Shadish, W. R., Jr., Cook, T. D., & Leviton, L. C. L. (1991). *Foundations of program evaluation*. Thousand Oaks CA: Sage.
- Springett, J. (2001). Participatory approaches to evaluation in health promotion. In I. Rootman, M. Goodstadt, B. Hyndman, D. V. McQueen, L. Potvin, J. Springett, & E. Ziglio (Eds.),

- Evaluation in health promotion: Principles and perspectives* (pp. 83–105). Copenhagen: WHO Regional Publications, European Series, No 92.
- Stake, R. E. (Ed.) (1975). *Evaluating the arts in education: A responsive approach*. Columbus Ohio: Merrill.
- Strauss, A., & Corbin, J. (1998). *Basics of qualitative research: Techniques and procedures for developing grounded theory*. Thousand Oaks, CA: Sage.
- Suchman, E. (1967). *Evaluative research*. New York: Russel Sage Foundation.
- Thurston, W. E., Graham, J. & Hatfield, J. (2003). Evaluability assessment a catalyst for program change and improvement. *Evaluation and the Health Profession*, 26, 206–221.
- Thurston, W. E., & Potvin, L. (2003). Evaluability assessment: A tool for incorporating evaluation in social change programs. *Evaluation*, 9, 453–469.
- Weber, M. (1952). The essentials of bureaucratic organization: An ideal-type construction. In R. K. Merton, A. P. Gray, B. Hockey, & H. Selvin (Eds.), *Reader in bureaucracy* (pp. 18, 21–22). New York: Free Press.
- Weiss, C. H. (1998). *Evaluation. Methods for studying programs and policy*, 2nd edition. Upper Saddle River NJ: Prentice Hall.
- Wholey, J. S., & Newcomer, K. E. (Eds.) (1989). *Improving government performance : evaluation strategies for strengthening public agencies and programs*. San Francisco: Jossey-Bass.