Evaluation

SAGE Publications (London,
Thousand Oaks and New Delhi)
DOI: 10.1177/1356389006064189
Vol 12(1): 56–76

Evaluation Pragmatics

CLAUDIO BEZZI
Evaluation Consultant, Italy

Evaluation is certainly based on social research, but each social research method and technique has its own limits and generates its own specific biases. This issue becomes critical at a fundamental evaluation point: the moment when one is required both to define the subject and mandate, and formulate the 'evaluative questions' which guide the whole activity. In this phase, it is important to look at every evaluation as a *local* issue, leaving the subject's definition to the (anthropological, organizational, professional) context, its grammar and lexicon. To be useful, an evaluation must focus on the context, i.e. on the 'texts' produced by the stakeholders and the meanings they give to these texts, thus helping them to contextually reduce evaluand ambiguity. In this perspective, evaluation is a learning organization's tool that helps to build social reality. Such a tool has its own mechanism, which is recognized by the *stakeholders* and makes possible both the intervention and the success of the programme.

KEYWORDS: evaluand semantics; evaluation context; evaluation methodology; evaluation pragmatics; learning organization

Objectives

This article seeks to underscore a few epistemological issues, which can be summarized as follows.

- An evaluation is not just a matter of techniques. As seen in the contemporary literature, the evolving international debate seems to take this assumption for granted. Yet, the real, 'factual' level of the European field continues to be *very* technical, most often without much sophistication.
- We need to clearly show *why* the predominance of a technicist approach is 'wrong', and that it leads us away from our evaluative objectives. This problem, which is far from settled, is very much 'European', in view of the link between the evaluation in the Old Continent and the European Union logic, with all of its rigidity.
- The fight against technicism is commonly labelled a 'constructivist' approach. This is not necessarily the case: thinking in methodological and epistemological terms, opposing undue simplifications, rampant reductivism

and logical shortcuts, serves any evaluation and its results. Although it might echo 'realistic' themes, this article is not concerned with the realism/constructivism dichotomy.

- This article presents the following core thesis: there are many reasons why we cannot clearly see the evaluand. The continuing debate upon the so-called objective clarification illustrates the difficulty in comprehending an evaluated programme in all of its complexity. Evaluators have attempted to solve this problem, here labelled the 'black box', in different ways that will be summarized.
- With regard to the black box problem, one of the most interesting recent proposals has been put forward by Pawson and Tilley. The relationship they postulate between contexts and mechanisms, called 'Realistic evaluation', is particularly useful in view of its systemic overview of that slice of social reality we call the evaluand. Nevertheless, while the contexts–mechanisms dyad constitutes the very fulcrum of realistic evaluation, today's debate often focuses on mechanisms alone, as if they were the evaluators' main concern. On the contrary, this article attempts to shift the focus on to contexts. Numerous findings of sociological, anthropological, linguistic and psychological research support an interpretation of the contexts as the real foundation of both the reality embodied by the evaluand, and the very capability to evaluate it. The current interdisciplinary reflection on this matter, briefly summarized here, seems to strongly support this theory.
- The highlighting of context has an important effect on the role of evaluation as a learning organization's tool. While it does not change the evaluation mission, nor redefine it in a more limited or specific fashion, such a context-oriented role can help us to solve some of the old continuing quandaries about the use of evaluations.
- Finally, this article seeks to illustrate that these issues are not a mere philosophical quest, but constitute a factual methodological approach, built on specific techniques and procedures.

Evaluation as Research and the Black Box

The Lack of Method

Evaluation, seen as a necessary and unavoidable technical process, and its possible failure, due to the almost certain systemic biases built into the method, often constitute an aporia.[1] An epistemological approach to this issue leads us to the following conclusions.

Evaluation is too often reduced to mere techniques. This approach, so-called 'operationalism' (Bridgman, 1964; Lundberg, 1939), is the capital sin in social and evaluative research. While technical intervention builds up information (as is partially inevitable), the two concepts of technique and information overlapping, without any conceptual distinction, can lead to a *possible* construction/interpretation of such information and the evaluand as well.

Techniques are both the *fundamenta* (foundations) and the original bases of social and evaluative research. Techniques produce findings, but what we find becomes a sort of reflection of the evaluand, a shadow of reality. There is no social (and evaluational) research technique, no matter how evolved, which is exempt from restraints. Questionnaires, cost–benefit analysis, contingent valuation, experimental approach, etc. are all useful or needed, but every one of them is subject to considerable limitations. This realization is one of the reasons behind the reflection on so-called *triangulation*, originally seen as a strategy to increase the validity of evaluation, overcoming the limitations typical of each technique.[2]

Nevertheless, the very concept of 'reality' is ambiguous and dangerous. At this stage, it seems as useless to offer a definition of reality as to protest any scientific or philosophical use of that term. A simple observation shall suffice: 'reality' is the plane in which evaluand, evaluator and his/her techniques coexist. Whatever the meaning attached to 'reality', first and foremost we partake in it, together with the data and all the information we seek while evaluating. Ourselves, our sponsor, evaluand, data, techniques and final evaluative judgment, everything lies on the same plane of 'reality' and of meaning. Therefore, the so-called 'findings' are merely the product – like others – of the same reality where our thoughts were born. This assumption is methodogically relevant, at a very factual level, because it is anchored on an explicit pragmatics of evaluation (analogous to pragmatics in linguistic, as discussed later). It refers to the existing relationship between data and those who generate them, demystifying the trend towards 'pretended sacredness' typical of operationalism;

Following this approach, we can reach only one conclusion: when we evaluate something, we do not observe the evaluand in its ontological core, but – on the contrary – we only observe what the stakeholders say, what they are able to account for, and what we are able to understand. Arguably, nothing is as important. The items evaluators talk about (the evaluand, its efficacy, programme theory, etc.) are *just texts.* The programme and its logic (including need analysis, problem tree, goal levels, evaluation mandate, objectives, evaluation development and its context); our techniques of every colour and description; the evaluation findings (whether numeric tables or recommendation sets); whatever we do and the contexts in which we do it, all belong to the evaluational domain as long as it can be expressed. In turn, various means of expression (reports, statements, programming plans, focus group analysis, data array) are conditioned by the way in which they were comprehended and interpreted by each stakeholder, including the evaluator, according to his/her formal, moral and normative restraints (what one 'can' say); personal conventions and choices often undeclared (what one 'wants' to say); and various skills (what one 'knows' how to say). Furthermore, we are all subject to a number of constraints, such as our cultural frame and individual psychology (and events like stomach aches, spousal quarrels of the previous day and so on, to be added to the 'moderators', as discussed by Dahler-Larsen, 2001).[3]

If we are in a hurry, if we have nothing to offer but a prepackaged set of techniques, if we ignore social research complexities, in all likelihood we will (inadvertently) adopt an operationalist approach, rushing towards the 'data', viewed as certain beyond discussion. Thus, we might embrace procedures, techniques and tools, deluding ourselves with *the illusion that techniques 'generate' an unveiling of reality.* This 'reality', perceived as unambiguous, is supposed to be waiting for

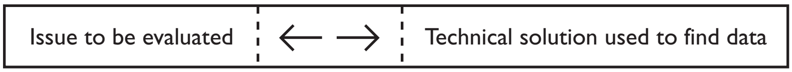

Figure 1. Evaluational Operationalism

the observer and his/her discovery (Figure 1 shows a simplified representation of this process).

But if we want to understand something about the problem at hand, we must develop a more complex process. In doing so, the key issue is the *preliminary analytical inquiry about what we are evaluating.* This analysis is indispensable in order to build a *programme theory* which avoids two serious pitfalls. On one hand the programme theory cannot be superimposed upon the theories formulated by the protagonists of the programme to be evaluated. In other terms, it cannot merely be an evaluator's construct which occurs when techniques overshadow the problem, as suggested by Dahler-Larsen, 2001: 332). On the other hand, a programme theory cannot be the figleaf of a stakeholder's 'theory'. We must keep in mind that such a text has been constructed by that stakeholder and the existence of other texts, striving to reach a *face validity* shared by a given context or organization. Then we can adapt appropriate operational procedures to the evaluand, as defined by our analysis (Bezzi, 2004).

This is the unavoidable frame in which we operate. This is not a constructivist ideology, nor a 'qualitative' as opposed to a 'quantitative' credo. The qualitative/quantitative (or constructivist/realistic) debate is unproductive and an obstacle to carrying out a more effective evaluation (Patton, 1998a). By necessity, evaluators are multidisciplinary professionals. Anthropology, semiotics and pragmatics inform us about the difficulty of defining 'reality' in an unambiguous and highly formalized way.[4]

The Black Box

Faced with this ambiguous reality, evaluators encounter their main concern: the black box issue, also known as the question 'what are we actually evaluating?'.

The black box is an apt metaphor for the impossibility/inability to 'see inside' the evaluand, its processes and the context which has generated it. Inside the black box, by definition, 'things' occur which are not well understood nor understandable, but from which results 'jump out'.

According to Stame (2004), we can describe the different points of view on the black box in the following four ways:

- The black box does not exist, because we can describe any level of the evaluand objectives, and their relationship to the results. This is the 'logical framework' perspective, which still enjoys an unfounded currency.
- There is a black box issue, due to the ignorance of decision-makers. The evaluator has a noble quest: to feed decision-makers with social theories so that they might improve their decisions, plans and programs (Chen, 1990).
- There is a black box issue. Even better, there are a number of black boxes, one for each stakeholder involved. The evaluator has the difficult task of

constructing a consensus among stakeholders, starting with the best stakeholder (Weiss, 1977).
- There is a black box issue, but we have to focus directly on what works. We have to build a 'change theory' based on the comprehension of mechanisms and context. This is the well-known realistic evaluation, proposed by Pawson and Tilley (1997). In this case, the evaluator has the enlightening task of revealing context and mechanisms to decision-makers and other stakeholders.

These relevant attempts to open the box are critical of easy 'technical' shortcuts. Pawson and Tilley's proposal deserves to be discussed first, because it directly impinges upon the role played by the context.

Focus on the Context

Where is the Emphasis?

According to this very important evaluative perspective, the debate within our community is too focused on mechanisms. There is a very simple reason for that: mechanisms are perceived as being the main contributors to the programme's performance. They form the basis for judging a programme as 'successful'. Context, on the other hand, is perceived as a given. But this is a positivist perspective. It is merely an assumption to think that this is 'real' and that we *can* understand this reality by utilizing the best tools (data gathering, surveys, etc.). In order to eschew a priori 'certainties', one should adopt an anthropological point of view, dealing from the beginning specifically with the context. Context is then defined as the belief set of a community (whether large – e.g. the whole population of a country – or small – e.g. the staff of a local social service agency). From this perspective, mechanisms do not lose their importance in evaluating a programme, but their understanding arises from the context in which they were born, and which illuminates them.

Context Semantics

The basic problem is semantic in nature: we observe and interpret the world, interacting in and with it, through a system of signs (first and foremost, language). This system is the product of historically stratified social convention and agreements. We communicate along conventional pathways, and equally conventional is the set of priorities which inform our communication – the ways in which we debate and the very objectives of our discourse, tools and techniques.

It will suffice to consider the most common and important aspects of our social experience. 'Quality', 'efficacy', 'transparency', 'fairness', together with the whole set of concepts are used daily. We avoid trouble by joining a school of thought, routinizing our evaluation behaviour or ignoring the problem altogether. While there is insufficient space to dwell on it here, this issue is an everyday matter. We easily recognize that these concepts are not objective and, in view of their fundamental approximation, we feel compelled to define them time and time again, in accordance with usage, scholarly discipline, etc.

Concepts are social constructs with a strong contextual meaning (Geertz, 1983). We do not deal with unmoveable, unchangeable 'things', but with socially agreed-upon concepts, in real situations (Valovirta, 2002).

Evaluative Problems and their Local Meanings

When an agency director asks to evaluate his/her programme or service, s/he presents an abstraction, not a factual product. Subject to a well-known semantic arbitrariness, this abstraction must be understood, clarified and univocally defined on the basis of its real use by the interested social parties (Wittgenstein, 1963).

Evaluating an employment policy does not mean evaluating the printed data (administrative and political statements), training courses, software transfers to various agencies, etc., which this programme utilizes. Rather, these important components (mission, training initiatives, etc.) can be observed only if *everyone* (or at least evaluator and sponsor) *has a clear idea of what is being talked about.* Then the programme becomes something else altogether. It morphs into political, economic, social and administrative goals and expresses the multiplicity of interests, needs and thoughts of those who have shaped its various stages. These interests and goals, their nature and proponents notwithstanding, are not simple 'pieces' of the evaluand that can be all reassembled into a whole using patience, guile and correct techniques, like a puzzle. On the contrary, these elements are produced by different opportunities, shaped by different stakeholders and detected with different tools and techniques. It is a puzzle composed of unmatched shards. The only alternative to irrelevance is to transform the evaluator's task into a metalinguistic operation, translating the various pieces, interpreted by stakeholder interaction, and creating a set of interests and goals the sense and meaning of which are shared by all. Then, and only then, can we deal with the explicit or implicit levels of social action underpinning the programme itself. The alternative would be to evaluate the programme by counting the planned vs offered courses, the number of certified attendees, or perhaps, pushing the envelope, the number of attendees who have found employment.

These elements are not trifles, but their understanding needs a continuous comparison between different aspects: the general framework that defines and signifies the various components (i.e. the employment policies of the time, etc.) and the local interpretation given to them, reworking the general framework.[5] This continuous process of signification, balancing 'general' and 'particular', 'theory' and 'facts', must start with the constructs generated by social actors and the meanings these actors attribute to them (Geertz, 1983).

The evaluator must understand the *evaluation demand* to which s/he responds. S/he must know *why* s/he is asked to evaluate something, what the objectives are, who the *stakeholders* are. Only then can one understand the programme's goals, both implicit and explicit (or any other issue to be evaluated), selecting among the multitude of possibilities the elements (indicators) that better characterize the problem at hand. Based upon these elements, the evaluator can initiate a research process (chosen from amongst many) capable of tackling the specific cognitive issue linked by the evaluation to decisional, administrative or organizational problems.

Opening the Black Box and Reconstruction of a Programme Theory

Now the issue has changed. The sequence is no longer 'I understand what you are talking about (a thing, let's say a table) and your problem (to know its weight), and I have the tool you need (a scale).' We must say instead: 'I cannot understand what you are talking about until I complete an in-depth analysis. I must clarify your unique point of view, here and now. Only then can I offer some (and not all) elements useful to furthering your understanding, without pretending to offer you "explanations".' Thus, our evaluation becomes an 'act of meaning construction' (De Ambrogio, 2003: 28).

Our misunderstanding is caused by the difficulties encountered in recognizing those mechanisms, because they belong to the thought domain, not to the realm of 'reality', and the realm of thought is a product of our culture: a social construction (see Figure 2).

World of Thoughts, Value Construction and Evaluation (Box 1)

Alberto Marradi (1994) – following Ogden-Richards and Popper – classes all elements of knowledge into three 'worlds' or 'domains':

- World 1 of Referents: what we think and talk about (objects as well as actions, events, etc.);
- World 2 of Thoughts: the domain of concepts and statements;
- World 3 of Language, or of signs, in general.

This triadic distinction and the interconnections between worlds are considered by Marradi as crucial elements with a special resonance for research methodology and indicator construction. The evaluand exists in World 1, the evaluator operates within World 2, and his/her tools are exclusively taken from World 3. The relationship between 'world', 'language' and 'science' is better expressed by B. L. Whorf (1956).

Semantic Context and its Expression (Box 2)

This 'world of thoughts' can be represented as a semantic context in which various actors contribute to define 'finite provinces of meaning' (Schütz, 1971: 230), which are contractual in nature and have apparently 'imprecise' features. Obviously, the imprecision is only apparent – otherwise we would need to assume again that a World 1 reality can be only univocally embodied in Worlds 2 and 3 – and it does not inhibit daily social life reproduction (Berger and Luckmann, 1966). Nevertheless, it constitutes an immediate and clear problem for the more evolved scientific and evaluational language. If the evaluator must express an evaluand value (World 1), s/he can only depend on recordable information belonging to World 3. But because this information expresses real people, in a natural language conditioned by various factors, such as culture, context, individual outlook, the spur of the moment, etc., it is evident that any evaluator or social researcher can only deal with a *text* using his/her interpretation, introducing his/her own set of understandings in the process.

Evaluand black box	
Epistemological aspects of evaluand	Methodological aspects of the evaluator → evaluand relationship
[1] The evaluand originated in the world of thoughts (Marradi, 1994), therefore ideas ' are subordinated to the values and culture of the various social subjects participating in evaluand construction.	[2] The common ground upon which we can reflect, discuss, act (and evaluate) is a semantic context. The 'world of thoughts' becomes the 'world of language' and acquires value insofar as it is *accounted for,*[a] *expressed, stated*.
[3] The evaluand upon which we can act, or the subject of our planning, organizational and evaluative operations, can only be the product of the main decision-makers' consensus. Therefore, we must ignore the 'betrayal' (Watzlawick, 1976; Bezzi, 2003: 261–2) or 'drift' from its original meanings, the breadth and awareness of which remain unknown.	[4] Any reconstruction of the Programme objectives, theory, mechanisms, etc. is bound to become a *new* construction, interpretation, *account*. It becomes a new 'betrayal', no matter how carefully we analyse its needs, context and mechanisms.
[5] Resorting to the *stakeholders* during the various declensions of a participative evaluation does not eliminate the problem. Doing so, we simply shift the evaluation plane. Any 'group thinking' developed through a participative approach allows the group members to re-construct a somewhat new evaluand and to co-construct and legitimize it.	

[a] 'Accounted' with the same meaning as used in ethnological methodology (Garfinkel, 1967). Some ethnomethodological aspects of evaluation are discussed by Paoletti (1998).

Figure 2. Evaluand Black Box

> Unless (or until) we can open the skull and observe the mind from without, all the material at our disposal is derived from inferences and personal recollections, their known unreliability notwithstanding. (Watzlawick et al., 1967: 35)[6]

'Data' per se do not change the situation, despite the love for them professed by so many evaluators involved with the evaluative version of 'monitoring' activities. After all, data are constructed by the same social actors and within the same research environment in which the evaluators evaluate according to priorities, agendas, modes and practices which are social and contingent in origin, even though reified. As a consequence, data share the same problems as textual information.

Both information and data are ways by which the context communicates with the evaluator and this interaction, like any other relationship, is affected by various distortions and interpretations. This is well known to pragmatically oriented linguists and anthropologists alike, whose task is to understand the context and meanings of the 'natives' (whoever they might be, including the policies and the actors evaluated by us). As indicated by Duranti (1997), anthropologists consider the conceptual constructs expressed by the members of a given culture as local theories to be explained.

Evaluators can benefit by referring to 'local theories' (echoing Merton's 'middle-range theory', 1968) and to conceptual constructs to be explained (coherently with Pawson's realistic setup). The *action* of communicating (information production) can only concern what is revealed, observed and recorded. A communication which is hidden, undisclosed (even when it is our fault) is as good as non-existent, because evaluators know and can use for their evaluations only the information expressed to and collected by them.

The Text Betrayal (Box 3)

Language theory has taught us that interpretation is always a betrayal. It is useless to expound again on well-established theories and findings about communication actors' different linguistic skills (Bloomfield, 1984; Labov, 1972); cultural differences between speech communities (Duranti, 1997); semantic ambiguity, both connotative and denotative, of our utterances and non-verbal expressions (Eco, 1976); translation problems raised by slang or foreign languages (Watzlawick, 1976); Freudian misunderstandings and lapsus (Freud, 1965); communication perceptual distortions (Carpenter and McLuhan, 1960); the affective dimension of communication and its understanding (Donohew and Sypher, 1988); gender differences and linguistic usage (Hall and Bucholtz, 1995), etc. References are legion, well beyond the scant emblematic pointers here offered.

Other socioanthropological betrayals are present as well. As has already been stated, the evaluand is a 'text', but as such it has been written by several different hands, at different points in time. Let's consider a generic development programme. In all likelihood, in the background there lurk rules, regulations and deadlines, both national and supranational (such as those imposed by the EU agenda), which have been negotiated elsewhere through a process similar to one described here. These rules and regulations are normally adapted to various local contexts (i.e. regional domains) by ill-defined choices made by local bureaucrats under the whip of political mandates and pressures. In their respective ways, both bureaucrats and politicians are collectors of interests, needs and demands expressed by multiple local actors, whose desires must be fulfilled with various degrees of urgency. The programme, formulated by so many different parties, is then written down using a formal bureaucratic legalese. This procedure slightly reconfigures the meaning, not least because of the subtle effects of its contrived lexicon and syntax. Subsequently, the recipients read the programme and agree to implement some of its parts. In doing so they rarely interact with the programme formulators, dealing instead with external consultants or other actors. The real process is much more layered and complex, but even this simplified version suffices to deny legitimacy to any textual univocity, revealing the continuing inevitable 'betrayal' of the programme's original intent (if there ever was only one).

In addition to the multiplicity of actors and their differentiation, there is the problem of time and its changing effects on a given context. In 1958, Bateson defined the similar process of 'schismogenesis' as a process of behavioural and communication differentiation, caused by individuals' cumulative interactions. There is not a stable, definitive 'reality' that can be univocally described by the actors who participate in it. Interaction, exchange and their ever changing results cause cognitive shifts among the very same protagonists (Lanzara, 1993), to the point of interpreting the same phenomena in completely different ways and building different realities. Nothing can be as fruitless as the quest for the ultimate original truth, no matter how the term is interpreted. 'What constitutes the context changes during the interaction and is interpreted, restricted or enlarged as appropriate' (Duranti, 1997: 284).[7]

Evaluation Pragmatics (Box 4)

We must logically conclude that absolute meanings cannot be grasped, but only reduced to 'local' meanings, as with late Wittgenstein, transforming the 'indexicality' issue (Levinson, 1983; Montague, 1974) into a hermeneutical problem (Schwandt, 1997; Widdershoven, 2001). Indexicality is defined as the property of linguistic expressions to establish a direct relationship with their respective meaning, which is contextually defined. This feature is made obvious by demonstrative pronouns such as 'that' and 'this', personal pronouns such as 'you' and 'I', and temporal determinants like 'now', 'tomorrow', etc. Several authors have suggested that the majority of verbal communication is indexical in nature and therefore fully understandable only when referred to its context.

> The wisdom of the interpretive turn in the human sciences is that we learn that the objects, events, language and meanings that comprise the spheres of human action are indexical. That is, they depend on context. Only through situated use in discursive practices or language games[8] do human actions acquire meaning. . . . that what is considered good evidence and sound reasoning in evaluation depends in part upon the societal game being played in which evaluation is an important activity. . . . Activities like reasoning in evaluation or justifying claims in evaluation acquire meaning from the fact that they are situated within (or indexed to) this 'language game'. (Schwandt, 1997: 69)

To follow a different path, we must raise our sights to *evaluation pragmatics*. At certain stages, evaluation semantics is inevitable and effective. It helps us to search for both the *meaning* of linguistic expressions beyond the moments in which they are uttered and the relationships between linguistic expressions and worldly objects. For example, what is the general meaning of evaluational tools and objects? What is a Regional Operative Programme? What is a 'realistic' evaluation? What is the meaning of a given sentence expressed by a focus group? Nevertheless, *evaluation pragmatics* takes us straight to the relationship between signs and speakers, linguistic expressions and those who use them. Pragmatics focuses instead on *purposes* and *contexts*, the very core of both theory-based and realist evaluation.

Thus emerges an interesting parallel between Pawson's concepts of 'context' and 'mechanism' and their analogue 'context' and 'performative utterances' in pragmatic theory (Austin, 1962), as well as ‘context’ and schizophrenic semiology in psychiatry (Watzlawick et al., 1967), with many other echoes in anthropology, psychology of knowledge theories, etc., which this article cannot adequately explore.

To summarize, evaluation means production and interpretation of clues. Interpretation includes inferential processes based on utterance format (semantic level) and context (pragmatic level). Context, in turn, resolves into three types of information: 1) the interpretation of recorded utterances; 2) the perception of the situation in which they were uttered; and 3) the speakers’ encyclopedic knowledge, especially in the interpreter’s case. ‘*Context* . . . can be more or less limiting, but in some fashion always determines the contingent situations’ (Watzlawick et al., 1967: 122).[9]

Participation as Pragmatic Strategy (Box 5)

The last step can be labelled ‘co-production’ and sharing. The choice of participative paths is too often motivated by ideological factors that do not concern us. What counts instead is the *instrumental* role of sharing within evaluational research. As we have seen, the evaluand is a social ‘object’ (a process, an organization, a programme) which can be known only through communication acts, mainly linguistic in nature. These acts are necessarily semantically ambiguous and unfathomable unless linked to their pragmatic contextual intentions. To this effect, the various actors’ interventions achieve a simple and methodology-oriented result: they share the *pragmatic use* of the pivotal evaluation concepts (Morissey, 1995). This concept sharing has two advantages:

- The evaluator can build (or have built) a semantic platform which gives meaning to his or her research design and future results, minimizing the overlap of meanings and interpretations offered by self, the sponsor, one or many actors.
- The actors can recognize themselves in such a platform, ascertaining its *face validity,* as Patton calls it, and pragmatically use it during the following evaluation phases. The end result avoids the well-known rejection response elicited by any evaluation centred on technicalities, imposed somehow upon the group or actors while remaining forever alien both in language and process.

In any event, it is a matter of promoting dialogue among and between actors and the evaluator.[10]

Evaluation in the Cultural Ocean

What kind of evaluation should we make in this cultural context? A less pompous one, perhaps. Certainly, a less positivist one. Is this possible? No simple answer is immediately available. Caution is needed, as well as a more multidisciplinary and receptive attitude towards social complexity.

Thus, we become less apprehensive about the truth, the ‘real’ reality, the actual mechanisms and so on. While our interest has not changed, we focus now on the version accounted for by our stakeholders (the ‘natives’ of the evaluated culture). The decisional context is what decision-makers are able to imagine. The operative context is what the professional culture, *there*, has built in several years of work, by *that* staff. That context is recognizable insofar as *they* are able to account for it. Our evaluated context becomes what we are able to recognize with the help of multiple research techniques, each subject to its own biases and limitations, according to the data provided by local informants.

Evaluation is not a truth provider, but a *declaration* of what we have established to approximate truth, in a way that the involved stakeholders found acceptable. In other words, in a participative setting (with or without real participation), evaluation has a role in helping stakeholders to describe the world (the context) as well as they can, directing them to discover the specific mechanisms they are able to define.

Evaluation becomes a path to organizational learning, which circumscribes a world. Not *the* world, but only *a* world, relevant to the stakeholders involved, and likely to be effective for evaluation purposes (Owen and Lambert, 1995; van der Knaap, 1995; van der Meer, 1999).

In reality, evaluators are like movie directors. They (more or less consciously) help to shape interaction and opinions (Patton, 1998b: 229–30). They facilitate the stakeholders' realization of their role within a *newly created* setting. Data (generated by the context), their analysis (determined by acquired social knowledge), and the final judgement are no more than one possibility made real by the evaluation process.

Therefore, our evaluation does not have meaning and value simply because of a final evaluative judgement or the application of relevant methodological procedures. Its merit lies instead in discovering unknown meanings, which help stakeholders to develop a new self-awareness, and in implementing new connections between people, actions and thoughts.

Evaluation as Social Construction of Assessed Reality, in Practice

How does this kind of evaluation work in practice? Two case studies conducted using this approach provide examples. Lack of space forces us to review only a few aspects of these cases, limiting us to an in-depth exploration of the evaluand reality, in order to show its contextual valence. Images will be used: in view of the intrinsic limits of *texts* (see box 3 in Figure 2 and its narrative) it seems particularly useful to use concept maps. This tool, while known, is still rarely exploited by evaluators (Knox, 1995; Yampolskaya et al., 2004).

Case Study 1: How do HMO Training Managers Define their Context?

A Health Maintenance Organization (HMO) Internal Training Service is involved in a new organizational process and asks for an evaluative system which would help both managers and operators. But what do they want to evaluate? What is the Internal Training Service and, specifically, what does 'training' mean in the particular case of this HMO? Within the Italian context, the problem presents several facets:

- improving the performances of doctors, nurses and other staff members, in order to provide the best treatment to patients;
- ensuring that all of the agency's personnel have the same chance to be trained, with the same access to career and salary improvements;
- adhering to relevant international principles, guidelines, etc., such as the ISO standard, Evidence-Based Medicine, and so on;
- reaching an agreement with the Italian Health Ministry to certify the courses, thus giving participants the possibility to accrue professional credits.

As one can imagine, there are various cognitive, trade union, bureaucratic and normative implications which inhibit an easy and quick definition of 'training' in this context. Furthermore the specific mission of the agency also plays an important role.

A training service manager had long elicited the help of a network of partners within the organization, using them as informants, training promoters and 'sensors' of training problems. These associates adequately represented a cross-section of all health professions: a psychiatrist, a veterinarian, a nurse, a radiologist, etc. up to 10 people in total. The manager wanted to involve them in the evaluation experience as well, in order to separate it from the service reorganization at large.

From the first meeting, it was obvious that the participants had completely different ideas about the concept of 'training', even though at the beginning everyone considered their own views to be complete and correct. The organization's quality controller (one of the attendees) thought about training in an international context and in terms that were both exogenous and formal, dominated by the concept of quality as an overall superstructure. The nurse and some physicians held a more scholastic view, equating professional training with acquisition of knowledge, but the more union-oriented professional could not forget that attendance at certified courses means career advancement. Finally the participants with administrative managerial roles were concerned with their own issues, for example new employees and their in-house training. Every one of them had his or her own idea. Each had merit, but the views expressed were not always compatible nor consistent.

In this typical case, 'training' is another multidimensional concept. How should it be evaluated? Who is responsible for establishing only one meaning? The service manager alone? The general director of the agency? The trade union? In a number of circumstances, one of these solutions might be the answer. But in our case study the service manager had a fairly understated role, and preferred to involve a large number of internal agency stakeholders.

During half a dozen meetings, the evaluator helped the group think about the specific meaning they wanted to attribute to key concepts (i.e. 'training'). The group had to define the concepts separating out their main dimensions, sub-dimensions and 'indicators', according to Lazarsfeld (1958). This way, each indicator is not the result of the available data, adapted ex post to the field's needs (as is often the evaluators' custom, see Fitz-Gibbon, 2002), but the conscious product of a group construction process. Only later can the available data be analysed according to this process. Naturally, the evaluator did not speak about concepts, dimensions or Lazarsfeld. He simply helped the group 'explore' the problem until a shared definition was reached, mixing, softening and integrating its different aspects. The service mandate was then defined as follows:

> To organize the training service in a way consistent with its system logic and quality standards, assuring both linkage and dialog with other systems and an accountability-oriented evaluation.

This definition might well be tortuous, but it found clear resonance within the group. It contains a few key concepts: 'to organize', 'training' (as yet not well defined), 'system logic' as well as the obvious 'evaluation'. Subsequently, each element was further analysed by the whole group. For instance, 'training' was defined as:

> Development of human resources, promoting knowledge, skills and abilities in a manner consistent with strategic directives, company's planning and knowledge levels to the service of health advancement.

This definition is quite rich with concepts deserving further scrutiny. The following discussion continued until the explored elements were operatively translated into indicators (not discussed here), which allowed the evaluator to build an internal evaluation system.

Figure 3 shows the tree of concepts, dimensions and subdimensions (the illustration is a simplification which must be interpreted according to its overall context). It was easy to reach the next step and build significant indicators, and

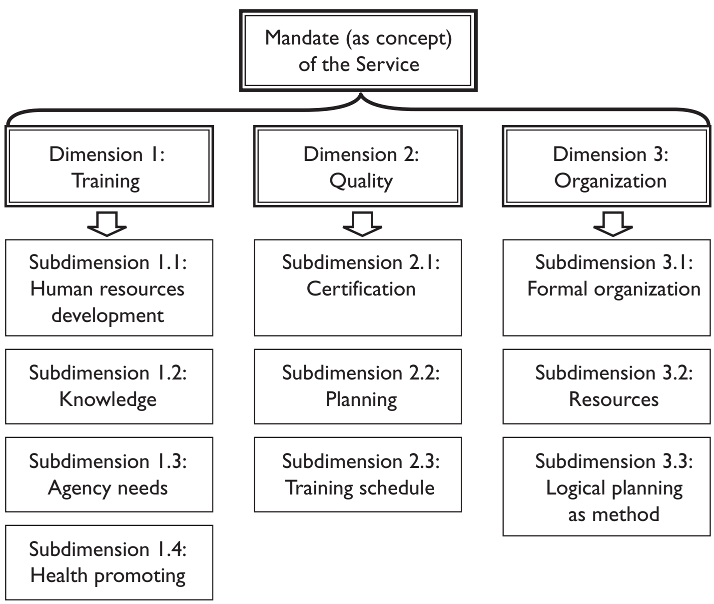

Figure 3. Training Evaluation Context: From Concept to Dimensions (Case Study 1)

then – at last – make them operative in an evaluative form. It is not important, here, to explain what the concept means (i.e. the 'mandate' in Figure 3), and why those dimensions were selected and not others. What counts is the following.

- The evaluator could have chosen to proceed with a common and routinized evaluation – for instance building a scale related to the usual training performances, or organizing several focus groups, etc. In this case, instead, he let the involved stakeholders, and – presumably – users, become aware of the complexity of their own context, and define it in their own way, starting from their organizational culture, the frame and its contents (i.e. the evaluand).
- Although it is possible to use a number of different procedures, our evaluator 'gambled on' the skills of the involved stakeholders, thus following a difficult path typically chosen by academic social researchers. It is not easy to explain 'Lazarsfeld's paradigm' to people outside the field of research and lead them through the whole process. In this case, it was possible to move from the core issue (what does *training* mean, exactly, in that specific context?) to its main components, and then to indicators.
- At the end, the result was not a mere list of indicators, but a new awareness of the relations existing between the agency's organization, the specific training service, informational needs, and evaluation. As a consequence, the evaluator and the training manager were able to build a compounded organizational and evaluative system.

Case Study 2: What is the Theory of the Programme?[11]

This is a very different context, characterized by greater complexity. We are focusing on a national Italian programme aiming to improve territorial services by implementing national employment policies. In this case, the ISI (Innovazione Servizi per l'Impiego, Employment Services Innovation) programme has the following specific goals.

- To standardize the information system on a national basis, with direct web access, in order to support any public or private service participating in regional networks;
- To organize and provide related technical assistance;
- To develop a method to support local public administrations.

Evaluating this programme presented a number of difficulties. First, the evaluator had to deal with the complexity of a national programme reflecting a wide variety of local approaches (the ISI programme spans 50 Italian provinces), each with its own peculiarities. The problem was further compounded by the ongoing evolution of the programme's method and the large number of staff members – subdivided between central headquarters and local offices. Third, the evaluator had to deal with the political relevance of the programme, the possibility of local resistance and so on.

When the programme started, the programme manager asked for an evaluation without specifying which kind of evaluation he had in mind, nor what the real evaluand was. The obvious answer would be the ISI programme. Unfortunately, it is almost impossible to evaluate such a complex, large, partially unclear programme. The evaluator has to understand exactly what should be evaluated and, in order to do so, has to make the involved stakeholders (the programme manager and his staff) aware of the evaluand itself. Ultimately, a programme is no more than what its stakeholders say it is. Experience shaped by working within large organizations and on vast programmes has shown that their complexity is not much greater than that of smaller programmes and organizations. In both cases, responsibilities are at stake, external factors – foreseeable to varying degrees – are bound to intervene (Dahler-Larsen's moderators), rules once considered clear and eternal must now be interpreted, informal and hidden rules have to be analysed. These problems are compounded by formal organization, goals, their schedules and budgets, and extant institutional relationships. How can we help the programme management? Only by helping them find their true position within this complex organizational spider web, highlighting potential management risks and the interaction of the various parts of the programme. In other words, we can help the team to reconstruct the programme theory. Programme theory is never rationalized in an univocal manner as if it were a law, and is never linear, not even in the manager's mind.

The evaluation task was exactly this: to help the programme staff understand, in no uncertain terms, what the programme was at that time, to build an evaluative proposal, and to update, afterwards, that representation in light of new developments.

There are several ways to help a team to rebuild a shared programme theory. Recently, Frans Leeuw (2003) has illustrated some interesting cases, the methodology of which could be distilled in the expression 'Search, Ask, Collect, Understand'. No technique per se can reconstruct someone's mental image. Therefore it is replaced by argumentation, with our intelligence for a 'tool'. Every time an evaluator fires up the argumentative skills of self and interlocutors alike, s/he probably is on track.

In a complete departure from the approach seen in the previous case study, the evaluator in this case constructed a conceptual map, eliciting from the start the help of programme management. Figure 4 shows the flowchart drawn during an in-depth discussion with the programme staff.

Without entering into irrelevant details:

- The programme staff had a number of different and unclear individual ideas about the programme. The evaluative opportunity allowed them to clarify in a unique way what they considered the programme to be (not what the programme 'really' *is*, because there is no such ontological truth).
- This activity, however, is not a simple clarification exercise. The programme thus defined reveals to ISI staff and evaluator alike its critical management components (indicated by arrows in Figure 4).
- Consequently, this reveals the key issues the evaluation has to address.

This 'map' was followed by a second one, illustrating the Programme Implementation Theory. Both maps were drawn by the evaluator in front of the management team, organizing the responses that emerged during a focus-group-like session. The maps were then revised by the evaluator, and submitted

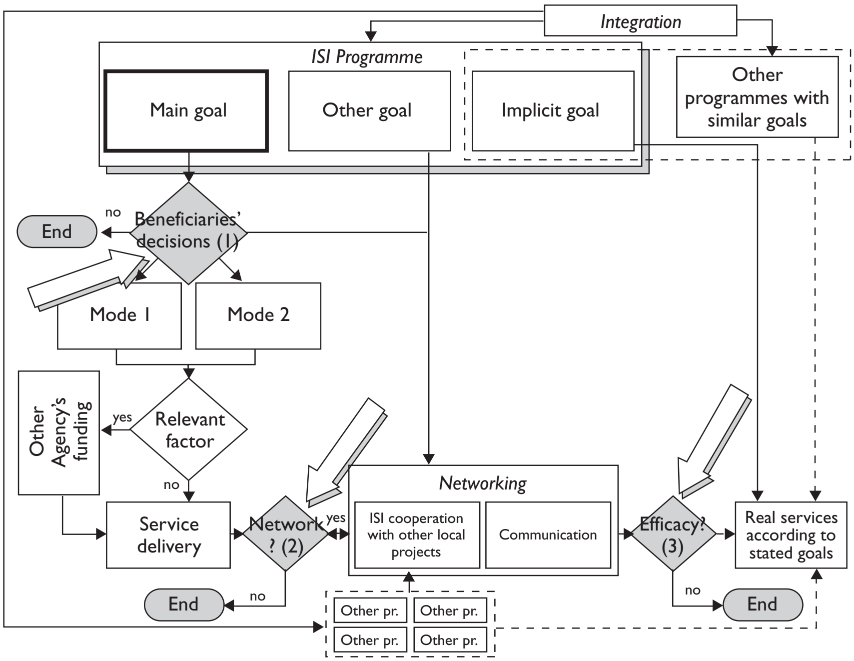

Figure 4. The ISI Programme Logic (Case Study 2; generic captions as original flowchart is confidential)

for the approval – in their final version – of the programme manager. These figures permitted the location of critical nodes, which defined the subsequent construction of an internal monitoring system and helped to establish the next performance evaluation goals. Clearly, we are not interested in the maps as such, but in the embodiment of a shared thought, forged through dialogue and argumentation. The evaluator did not present abstractions, looking instead for the programme theory within the specific context of that particular community of practices made up of the programme staff.

Back to the Context

This article focuses on the initial steps of the evaluation process, during which the evaluand has to be defined in a participative way, with a local grammar and lexicon. Because there is not a pre-existing and fixed reality, unambiguous evidence or a pseudo-divine revelation, we have to search for a new brand of 'truth' to be established by the involved stakeholders to the best of their knowledge and ability.

Evaluation does not reveal any truth, nor establish any findings predating the evaluation itself. Evaluation builds the reality or, in simpler terms, builds the evaluand. It is not important to make final conclusions regarding the participative question (whether participation makes for a better evaluation). Evaluation may be participative or not. If it is so, the evaluand is built in a way that stakeholders can appreciate, with a grammar and a lexicon familiar to the potential evaluation users. What matters is focusing the evaluation on the context and its actors, asking the fundamental questions: What do they say? What do they imagine? What are their values? What are they looking for?

The two case studies presented here briefly retraced some initial moments of dialogue and argumentation by relevant members of both programmes. In those situations, the evaluator did not seek an ‘objective’ truth, but just the ‘truth’ – surrounding the evaluand – held by the stakeholder, because it is useful to managing real processes and open to the unexpected.

Because of this author’s belief that the true engine of evaluation is represented by ‘evaluative research’, little credence should be given to the most extreme constructivist shifts. We need ample methodological reflection, which exceeds the boundaries of this article. Our literature offers a plethora of texts dedicated to research methods and techniques centred on discourse, texts, local cultures and programme theory reconstruction (most recently: Bezzi, 2005; Christie and Alkin, 2003; Christie and Rose, 2003; Dart and Davies, 2003). For now a few general methodological reflections should suffice:

- Although the technique shopping cart is loaded with attractive gadgets, whose number grows daily, we must beware the technicism denounced by this article. The evaluator’s soul cannot be saved by operationalist sophistry, no matter how interactive and colourful the software might be.
- By the same token, anti-technicism cannot be a justification for methodological approximation nor antiscientific crusades, as implied by some constructivist echoes (Guba and Lincoln, 1989; criticized by Kushner, 1996). If evaluation must be dialogue and argumentation, the need for a rigorous methodology and validated results is even greater, exactly because ‘dialogue’, ‘reality as a text’ and participation are by their own nature undetermined, subject to interpretation and mediated through sharing.
- The challenge – methodological rigour vs interpretation – can be met linking our evaluational work to contexts, local cultures (not to be confused with their territorially delimited variety), linguistic games and local uses of signs and codes. This approach alone can pay due attention to those moderators which otherwise undermine any programme theory.
- At the heart of the evaluation effort lies a new awareness about context complexity encountered by the sponsor or the manager of a given programme; the construction, together with the relevant stakeholders or the programme team, of a shared representation of the evaluand; and especially the sharing of a contextually valid lexicon and a grammar. Through this approach, it is possible to achieve two remarkable results. First, we understand what the evaluand *really* is (at least in that context, for those stakeholders), thus making possible the conduct of evaluation research which is adequate, pertinent and valid; and second, the interlocutors build a new awareness, which will help them to better manage the programme and to share the evaluation process.

Notes

1. This is not an article against techniques and the primacy of a rigorous methodological approach to evaluation research. As a methodologist, I am simply cautious about brazenly used evaluation techniques (see Bezzi, 2003). This article is focused on a quiet realization: techniques aren't the road to heaven.
2. See Mark and Shotland (1987). It should be noted how, in less than 10 years, the debate about Multiple (or 'Mixed') Methods had freed itself from the validity problem, in order to encompass a much larger scope, albeit with seemingly limited success: Greene and Caracelli (1997).
3. Within the evaluation field, the cultural dimension I am discussing is well defined by Hyatt and Simons (1999), who present several real case scenarios.
4. Nevertheless, browsing the annals of *Evaluation, American Journal of Evaluation, Rassegna Italiana di Valutazione*, and other evaluation journals reveals a surprising paucity of anthropological, psychological and semiotic perspectives in evaluation.
5. Macro (i.e. the programme to be implemented) and micro level (its comprehension and social reattribution of meaning, which retroactively affects the programme itself) interaction is a crucial element recently revisited by Palumbo (2001), along Max Weber's lines.
6. Retranslated from the Italian edition of this work.
7. Retranslated from the Italian edition of this work.
8. The language game concept is further developed, and to good effect, by Wittgenstein.
9. Retranslated from the Italian edition of this work.
10. The role of dialogue in evaluation was the main focus of an entire issue (7[2]) of *Evaluation,* edited by Tineke A. Abma. See also Christie and Rose, 2003.
11. A paper on this case study was presented to the VII Meeting of the Italian Evaluation Society in Milan, 2004. I thank Italia Lavoro SpA, the source of this material, for allowing its dissemination.

References

Austin, John Langshow (1962) *How to Do Things with Words.* Oxford: Oxford University Press.

Bateson, Gregory (1958) *Naven: A Survey of the Problem Suggested by a Composite Picture of the Culture of a New Guinea Tribe Drawn from Three Points of View*, 2nd edn. Stanford, CA: Stanford University Press; 1st publ. 1938.

Berger, Peter L. and Thomas Luckmann (1966) *The Social Construction of Reality: A Treatise in the Sociology of Knowledge.* New York: Anchor Books.

Bezzi, Claudio (2003) *Il disegno della ricerca valutativa*, rev. edn. Milan: Franco Angeli.

Bezzi, Claudio (2004) 'Cosa valutiamo, esattamente?', *Rassegna Italiana di Valutazione* 27: 49–73.

Bezzi, Claudio (2005) 'Rendiamo dinamica la SWOT', *Rassegna Italiana di Valutazione* 31: 7–15.

Bloomfield, Leonard (1984) *Language.* Chicago, IL: University of Chicago Press; 1st publ. 1935.

Bridgman, Percy Williams (1964) *The Nature of Physical Theory: Published on the Louis Clark Vanuxem Foundation.* Hoboken, NJ: Wiley.

Carpenter, Edmund and Marshall McLuhan (1960) 'Acoustic Space', in E. Carpenter and M. McLuhan (eds) *Explorations in Communications: An Anthology*, pp. 65–70. Boston, MA: Beacon Press.

Chen, Huey Tsyh (1990) *Theory-Driven Evaluations.* Newbury Park, CA: SAGE.

Christie, Christina A. and Marvin C. Alkin (2003) 'The User-Oriented Evaluator's Role in Formulating a Program Theory: Using a Theory-Driven Approach', *American Journal of Evaluation* 24(3): 373–85.

Christie, Christina A. and Mike Rose (2003) 'Learning about Evaluation through Dialogue: Lesson from an Informal Discussion Group', *American Journal of Evaluation* 24(2): 235–43.

Dahler-Larsen, Peter (2001) 'From Programme Theory to Constructivism: On Tragic, Magic and Competing Programmes', *Evaluation* 7(3): 331–49.

Dart, Jessica and Rick Davies (2003) 'A Dialogical, Story-Based Evaluation Tool: The Most Significant Change Technique', *American Journal of Evaluation* 24(2): 137–55.

De Ambrogio, Ugo (2003) *Valutare gli interventi e le politiche sociali.* Rome: Carocci.

Donohew, Lewis and Howard E. Sypher (1988) *Communication, Social Cognition and Affect*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Duranti, Alessandro (1997) *Linguistic Anthropology*. Cambridge: Cambridge University Press.

Eco, Umberto (1976) *A Theory of Semiotics*. Bloomington: Indiana University Press.

Fitz-Gibbon, Carol (2002) 'Evaluation in an Age of Indicators: Challenges for Public Sector Management', *Evaluation* 8(1): 140–8.

Freud, Sigmund (1965) *The Psychopathology of Everyday Life*. New York: W.W. Norton & Co.

Garfinkel, Harold (1967) *Studies in Ethnomethodology.* Englewood Cliffs, NJ: Prentice-Hall.

Geertz, Clifford (1983) *Local Knowledge: Further Essays in Interpretative Anthropology.* New York: Basic Books.

Greene, Jennifer C. and Valerie J. Caracelli, eds (1997) 'Advanced in Mixed-Method Evaluation: The Challenger and Benefits of Integrating Diverse Paradigms', *New Directions for Evaluation* 74.

Guba, Egon G. and Yvonna S. Lincoln (1989) *Fourth Generation Evaluation.* Newbury Park, CA: SAGE.

Hall, Kira and Mary Bucholtz, eds (1995) *Gender Articulated: Language and the Socially Constructed Self.* New York: Routledge.

Hyatt, Jenny and Helen Simons (1999) 'Cultural Codes Who Holds the Key? The Concept and Conduct of Evaluation in Central and Eastern Europe', *Evaluation* 5(1): 23–41.

Knox, Colin (1995) 'Concept Mapping in Policy Evaluation: A Research Review of Community Relations in Northern Ireland', *Evaluation* 1(1): 65–79.

Kushner, Saville (1996) 'The Limits of Constructivism in Evaluation', *Evaluation* 2(2): 189–200.

Labov, William (1972) *Sociolinguistic Patterns.* Philadelphia: University of Pennsylvania Press.

Lanzara, Giovan Francesco (1993) *Capacità negativa: Competenza progettuale e modelli di intervento nelle organizzazioni.* Bologna: Il Mulino.

Lazarsfeld, Paul Felix (1958) 'Evidence and Inference in Social Research', *Dedalus* LXXXVII: 99–109.

Leeuw, Frans L. (2003) 'Reconstructing Program Theories: Methods Available and Problems to be Solved', *American Journal of Evaluation* 24(1): 5–20.

Levinson, Stephen C. (1983) *Pragmatics.* Cambridge: Cambridge University Press.

Lundberg, George A. (1939) *Foundations in Sociology*. New York: The MacMillan Company.

Mark, Melvin M. and R. Lance Shotland, eds (1987) 'Multiple Methods in Program Evaluation', *New Directions for Program Evaluation* 35.

Marradi, Alberto (1994) 'Referenti, pensiero e linguaggio: una questione rilevante per gli indicatori', *Sociologia e ricerca sociale* 43: 137–207.

Merton, Robert K. (1968) *Social Theory and Social Structure.* New York: Free Press.
Montague, Richard (1974) 'Pragmatics', in Richard Thomason (ed.) *Formal Philosophy*. New Haven, CT: Yale University Press.
Morissey, Orla (1995) 'Shifting Paradigms: Discorse Analysis as an Evaluation Approach for Technology Assessment', *Evaluation* 1(2): 217–35.
Owen, John M. and Faye C. Lambert (1995) 'Roles for Evaluation in Learning Organizations', *Evaluation* 1(2): 000–000.
Palumbo, Mauro (2001) 'Valutazione e ricerca sociale', *Rassegna Italiana di Valutazione* 24: 29–48.
Paoletti, Isabella (1998) 'Un approccio interpretativo alla valutazione: il contributo dell' etnometodologia', *Rassegna Italiana di Valutazione* 10, www.valutazioneitaliana.it
Patton, Michael Quinn (1998a) *Utilization-Focused Evaluation*, 3rd edn. Beverly Hills, CA: SAGE.
Patton, Michael Quinn (1998b) 'Discovering Process Use', *Evaluation* 4(2): 225–33.
Pawson, Ray and Nick Tilley (1997) *Realistic Evaluation.* London: SAGE.
Schütz, Alfred (1971) *Collected Papers.* The Hague: Martinus Nijhoff.
Schwandt, Thomas A. (1997) 'Evaluation as Practical Hermeneutics', *Evaluation* 3(1): 69–83.
Stame, Nicoletta (2004) 'Theory-Based Evaluation and Types of Complexity', *Evaluation* 10(1): 58–76.
Valovirta, Ville (2002) 'Evaluation Utilization as Argumentation', *Evaluation* 8(1): 60–80.
van der Knaap, Peter (1995) 'Policy Evaluation and Learning', *Evaluation* 1(2): 189–216.
van der Meer, Frans-Bauke (1999) 'Evaluation and the Social Construction of Impacts', *Evaluation* 5(4): 387–406.
Watzlawick, Paul (1976) *How Real is Real? Confusion, Disinformation, Communication.* New York: Random House.
Watzlawick, Paul, Janet Helmick Beavin and Don D. Jackson (1967) *Pragmatics of Human Communication: A Study of Interactional Patterns, Pathologies, and Paradoxes.* New York: W.W. Norton & Co.
Weiss, Carol (1977) 'Theory-Based Evaluation: Past, Present and Future', *New Directions for Evaluation* 76.
Whorf, Benjamin Lee (1956) *Language, Thought, and Reality: Selected Writings.* Cambridge, MA: MIT Press.
Widdershoven, Guy A. M. (2001) 'Dialogue in Evaluation: A Hermeneutic Perspective', *Evaluation* 7(2): 253–63.
Wittgenstein, Ludwig (1963) *Philosophical Investigations.* New York: Macmillan; original edn *Philosophische Untersuchungen*, Oxford: Basil Blackwell, 1953.
Yampolskaya, Svetlana, Teresa M. Nesman, Mario Hernandez and Diane Koch (2004) 'Using Concept Mapping to Develop a Logic Model and Articulate a Program Theory: A Case Example', *American Journal of Evaluation* 25(2): 191–207.

CLAUDIO BEZZI is an evaluation consultant, specializing in the epistemological and methodological aspects of evaluative research. He was one of the founders of the *Associazione Italiana di Valutazione*, and in 1996 launched the *Rassegna Italiana di Valutazione*, an evaluation quarterly he is still director of today. He has published several articles and books on methodology and evaluation, including the textbook *Il disegno della ricerca valutativa*. Please address correspondence to: via della Zolfara – Cenerente, 10, 06070 Perugia, Italy. [email: bezzi@valutazione.it]