

American Journal of Evaluation

<http://aje.sagepub.com>

Evaluators' Decision Making: The Relationship Between Theory, Practice, and Experience

Claire Tourmen

American Journal of Evaluation 2009; 30; 7

DOI: 10.1177/1098214008327602

The online version of this article can be found at:
<http://aje.sagepub.com/cgi/content/abstract/30/1/7>

Published by:



<http://www.sagepublications.com>

On behalf of:

American Evaluation Association

Additional services and information for *American Journal of Evaluation* can be found at:

Email Alerts: <http://aje.sagepub.com/cgi/alerts>

Subscriptions: <http://aje.sagepub.com/subscriptions>

Reprints: <http://www.sagepub.com/journalsReprints.nav>

Permissions: <http://www.sagepub.com/journalsPermissions.nav>

Citations <http://aje.sagepub.com/cgi/content/refs/30/1/7>



Evaluators' Decision Making

The Relationship Between Theory, Practice, and Experience

Claire Tourmen

National Institute of Higher Education in Agronomy, Dijon, France

How do evaluation practitioners make choices when they evaluate a program? What function do evaluation theories play in practice? In this article, I report on an exploratory study that examined evaluation practices in France. The research began with observations of practitioners' activities, with a particular focus on the phases of evaluation design. The purpose of the study was to examine the logic underlying their choices and the role that evaluation theory played in those choices. In this study, I used activity theory and ergonomics methods. I highlight various ways in which working evaluators make evaluation choices in real situations, explained by different sets of conceptual resources they hold related to their level of experience. Finally, I show how this research contributes toward clarifying the role of theories in practice and enriching an evaluation theory

Keywords: *evaluation theory; evaluation practice; situated decision making; skills; training*

How do evaluation practitioners make choices when they evaluate a program? Why would they, for example, prefer one method to another? Do they refer to evaluation theories? In other words, how is evaluation practiced and how can it be taught? There is an abundant literature aimed at theorizing and prescribing evaluation practice. Many authors insist on the importance of understanding evaluation "theories" to know how to evaluate (Donaldson & Lipsey, 2008). Yet the exact function that theories play in evaluation practice is not well known. Donaldson and Lipsey (2008) contend that: "the nature and role of theory in evaluation is often a contentious matter" (p. 57). Is it enough to know a lot about evaluation theories? What role does experience play for evaluators?

Recent studies have explored these questions. Schwandt (2007) and Donaldson and Lipsey (2008) mention the "practical knowledge" that evaluators seem to use in their everyday activity. The way experienced evaluators solve the problems they face in real situations could thus be an excellent basis for learning more about the logic of evaluative activity. Do evaluation practitioners also have their "tricks of the trade?" There is, however, a problem in this approach, in that practical knowledge is embodied in activity flow and often remains implicit. How can it be characterized, described, and analyzed?

Author's note: This article is based on a PhD obtained from the University of Grenoble II (France) and co-financed by a private firm in program evaluation, Eureval, 13 bis place Jules Ferry, 69006 Lyon, France. I am grateful to Gérard Figari (University of Grenoble II) and Patrick Mayen (Enesad) for having directed this research, I am also grateful to Jacques Toulemonde and Eric Monnier (Eureval) for having co-financed it. I am finally grateful to William Shadish, who encouraged me to publish in AJE and kindly helped me to do it, and to Jody Fitzpatrick for her help and support. Please address correspondence to Claire Tourmen, National Institute of Higher Education in Agronomy, 26 Bd du Docteur Petit Jean, 21000 Dijon, France; phone: (+33) 380-772328; e-mail: claire.tourmen@educagri.fr.

In this article, I report on an explanatory study that examined evaluation practices in France. The study began with the observation of evaluation practitioners' activity, to study the logic of that activity and examine its relation to evaluation theories. I used activity theories and ergonomics methods to observe evaluators' practices, to "make the tacit explicit" (Schön, 1983) and possible to study systematically. The results are based on a sample of 24 evaluation practitioners working in France and at the European Commission. The observations were focused on evaluation design phases.

In this article, I highlight the various ways that evaluators make evaluation choices in real situations. These choices can be explained by different sets of "conceptual resources" that evaluators use when making decisions. In turn, the conceptual resources an evaluator has is related to his or her level of experience. The results of this study show that experienced practitioners have built up pragmatic knowledge that is different, but not completely disconnected from formal theoretical knowledge in evaluation. Bringing their practical knowledge to life makes its description possible and is a source of information for training evaluators and enriching descriptions of evaluators' skills. Finally, the contribution of this research toward evaluation theory is discussed, in the tradition of Shadish, Cook, and Leviton (1991).

Evaluators' Choices in Question

Evaluating is a common and daily activity (Stufflebeam, 1980), yet program evaluation is a specialty that is becoming more and more professionalized. Since the 1960s in North America and the 1980s in Europe, public administrations and governments have increased their demands for studies called "evaluations" (Furubo, Rist, & Sandhahl, 2002). In response, numerous actors in many countries—researchers, civil servants, consultants—have specialized in program evaluation. The emergence of practices called evaluation has resulted, encouraged by the production of an abundant literature—scientific and professional—offering conceptual definitions, prescribed tasks, and ways and means to evaluate. Professional groups have also emerged in several countries (according to Donaldson and Christie, 2006, there are more than 50 evaluation societies throughout the world). Their members question their own know-how and identity, and the professional nature of their work (see Worthen, 1994).

This article will examine evaluator's decision making by studying the choices of evaluation practitioners in France. In 1990, the French government defined policy evaluation as follows (22 January 1990 decree): "Evaluating a policy means analyzing whether the legal, administrative and financial means implemented produce the desired effects of the policy and enable its objectives to be achieved" (first article). Policy and program evaluation has increased since then in French administrations, also encouraged by the European Commission. In 2006, the French Evaluation Society had 250 members from public and private organizations. They have built their expertise on the evaluation of different public policies, programs and projects, in various actor systems, for different goals and in diverse ways (using varying degrees of collaboration, different methodologies, etc.). What kind of expertise have they built to manage such tasks?

Taxonomies of Skills

To describe evaluators' specific know-how, taxonomies of evaluation skills have been proposed by researchers and practitioners in the United States (King, Stevahn, Ghery, & Minnema, 2001; Mertens, 1994; Torres, Preskill, & Piontek, 1996), Canada (Kishchuck & Long, 1997; McGuire, Zorzi, & Perrin, 2002), and Australia (English, 2002). They describe

skills with long lists of knowledge (such as “research in social sciences,” “project management,” etc.), know-how (such as “analyzing data”), and tools and “social capacities” as general as “communication” (King et al., 2001) or the “ability to listen” (McGuire et al., 2002).

These authors have emphasized the fact that evaluators should know about evaluation concepts, methods, and tools to inform the decisions they are faced with throughout an evaluation. McGuire and colleagues (2002) assert, for instance, that evaluators “must select the most appropriate tools for a particular evaluation, in view of the evaluation context and the objectives” (p. 5), before going further. These studies have triggered debates on the way skills could be defined, the exact nature and role of knowledge of theory and evaluation concepts, and the possibility of reaching agreement on skills, given the wide range of evaluators' roles, approaches, and contexts. Moreover, these taxonomies do not offer insight into the choices faced in real situations nor into the way in which formal knowledge is or is not used effectively. Furthermore, what kind of knowledge is concerned?

Knowledge for Evaluation: Social Science Theories, Evaluation Theories, and Evaluation Theory

To clarify the debates, the three main kinds of knowledge encompassed by the word *theory* in the field of evaluation will be discussed. In a first sense of the word, authors refer to plural *social science theories* about the objects to be evaluated, that is, programs and the way they could be analyzed, drawing from, for instance, the fields of economy, political science, and sociology (see Chen, 1990).

In a second sense of the word, authors refer to plural *evaluation theories*, which are normative and prescriptive theories, competing models and approaches in program evaluation. The literature provides many directives on how people could or should evaluate. Numerous studies contribute toward defining a model way of designing and conducting good program evaluation. As Fitzpatrick, Christie, and Mark (2008) say, “Evaluators read and write about evaluation theories, approaches, and models—how evaluation *should* be practiced” (chapter 14). For instance, one could distinguish “experimental evaluation” from “utilization-focused evaluation” (Patton, 1986), also different from “empowerment evaluation” (Fetterman, Kaftarian, & Wandersman, 1996) and “theory-driven evaluation” (Chen, 1990), and so on.

In a third sense of the word, authors refer to *an evaluation theory*, which would be a descriptive theory about the way evaluations happen and produce different kinds of results and effects in given situations. As Shadish and colleagues (1991) affirmed: “the ideal (never achievable) evaluation theory would describe and justify why certain evaluation practices lead to particular kinds of results across situations that evaluators confront” (p. 31). As Fitzpatrick and colleagues (2008) stated: “We also read and write about the results of our studies and the methods we use to obtain those results” (chapter 14). What kinds of roles might these theories play in evaluation practice and in evaluators' choices?

Choices Could be Related to Evaluation Theories

The first role of knowledge in evaluation could be defined as a prescriptive role. According to several studies, evaluating always consists of two types of action: (a) comparing data on an object (such as a program) to standards (i.e., elements of reference stipulating how the object could or should be) to attribute value to the object and (b) rebuilding both the data on an object and the standards to make the comparison possible (Barbier, 1985; Bonniol, 1981; Stufflebeam, 1980; Tourmen & Figari, 2006). The latter action sometimes remains implicit (as in everyday

judgments, see Stufflebeam, 1980), and sometimes becomes explicit, as is often the case in program evaluation (through the actual building of questions, criteria, indicators and standards discussed and chosen during the initial evaluation steps).

The main difficulty stems from the fact that there is not one way to carry out these two actions but many: evaluators can try to achieve different goals and success criteria, and can do so in various ways, by using different degrees of collaboration with public partners and citizens and different data collection and analysis tools. These different ways of evaluating a program have been described and prescribed in what are called evaluation theories (Alkin & Christie, 2004), in the second sense of the word (see above). Alkin (2004) contends that evaluation theories offer a set of prescriptions, “prohibitions” and “guiding frameworks,” that specify “what a good or proper evaluation is” and “how evaluation should be done.” These theories have been grouped together in competing approaches, according to their similarities and differences (Alkin & Christie, 2004). They have also been called “generations” or “paradigms” of evaluation (Guba & Lincoln, 1989; Patton, 1986; Rodrigues, 2002). They are said to differ on the level of stakeholders’ involvement and the method proclivity (Christie, 2003). Thus, there are different and opposing ways of designing and conducting a program evaluation. We could therefore posit that evaluators’ choices could be related to a paradigm or a certain theory proclivity, and that evaluators just have to follow them step by step.

The Choices Could Also be Related to an Analysis of Real Situations

Apart from these differences because of theoretical stances, evaluators’ choices may also depend on which course of action is deemed most appropriate to the situation. Nadeau (1988) notes that “Stone (1984) asserts that excellence in program evaluation is based on the choice and use of methods that are relevant to contexts” (p. 48). Hadji (1989) also says that “Evaluators’ skills appear in their choices If you do not consider why and for whom you evaluate, technical questions turn out to be false questions” (p. 56). Recent studies try to go back to evaluators’ practices and to explore the reasons for their choices in relation to “contextual factors” (Fitzpatrick, 2004). These authors try to explore the complexity of real activities: “Evaluation practice, as any professional practice, is concerned with subtleties, nuances, or larger shades of difference in how evaluators *behave* during a study—what they consciously or unconsciously choose to do” (Fitzpatrick et al., 2008, emphasis added, chapter 14). They question the existence of “practical knowledge,” which they consider to be different from the formal knowledge of evaluation theories (in the second sense of the word, that is, various prescribed approaches).

Hence, another role of formal knowledge in evaluation could be to help evaluators to interpret the situations they are in and to make the most appropriate choices. Social science theories could help them to analyze the programs to be evaluated (see *Theory-Driven Evaluations* by Chen, 1990) and an evaluation theory could guide their choices. According to Shadish and colleagues (1991), “Evaluation theory tells us when, where, and why some methods should be applied and others not Evaluation theories are like military strategy and tactics; methods are like military weapons and logistics The good evaluator needs theories for the same reasons in choosing and deploying methods” (p. 34).

Shadish and colleagues (1991) nevertheless contend that this kind of evaluation theory is far from being achieved, and that evaluation practitioners are guided by their *own theories*. According to these authors, evaluators’ choices relate to their *own implicit theories*: “No method is appropriate always and everywhere All evaluation practitioners are nascent evaluation theorists. They think about what they are doing, make considered judgments about

which methods to use in each situation, weigh advantages and disadvantages of choices they face, and learn from successes and failures in their past evaluations” (p. 35).

The question is thus: how do practitioners evaluate a program? Do evaluators' choices depend on the evaluation theories they prefer? Do they depend on social science theories or parts of a formal theory of evaluation? Or do they depend on their own, implicit theories? What role exactly does formal knowledge play in real activity? Does experience bring specific knowledge and is it different from formal knowledge in evaluation? My research has explored these questions further by investigating evaluators' logic and decision making in context through work analysis theories and methods.

Work Analysis

A Theoretical Framework to Study Evaluation Practice

Following the stream of recent studies, I decided to change the perspective. I wanted to study what practitioners do and why they do so, but not how that relates to what they are believed or supposed to do. I wanted to go from practice to theory: above all trying to understand what people's concerns are, what terminology they use, and what reasoning they use, to apply this knowledge to a discussion of the evaluation literature, particularly the role of formal knowledge in evaluators' decision making.

I chose a theoretical model of activity to study the logic of situated decision making: the scheme model developed by Piaget (1955) and Vergnaud (1989, 1990). This model is based on the theories of activity (Vygotsky, 1934/1962), ergonomics, psychology of work and development (Chi, Glaser, & Farr, 1988), and professional didactics (Pastré, Mayen, & Vergnaud, 2006). Ergonomics studies have shown that real activity goes further than the prescribed tasks. Activity is organized at a conceptual level by a set of resources that enable people to adapt their actions to different situations with varying degrees of complexity. Situations surround and influence people's activities (Bruner, 1990). The resources that people use to act in/on a situation are developed and evolve through experience. Studies on novices and experts have shown that experts become more able to anticipate (Chi et al., 1988) and take into account the context of situations when making decisions (Dreyfus, Dreyfus, & Athanasiou, 1986). According to Vergnaud (2001), this can be explained by the development of “conceptual resources.” These conceptual resources can be described as *pragmatic concepts*, *pragmatic knowledge* about situations, and *action rules*, in other words, different forms of “conceptualization” of situations. In this way, studies of the psychology of work have shown that different levels of activity can be explained by different ways of conceptualizing situations built through experience.

I therefore sought to (a) understand whether the activities of practicing evaluators (with varying levels of experience) are organized by different ways of conceptualizing situations and (b) put forward a hypothesis on how these different ways of conceptualizing situations are developed. Finally, this study seeks to examine the influence of formal knowledge in evaluation activity.

A Work Analysis to Observe and Analyze Evaluation Practitioners' Activities

Inspired by classical methodologies in ergonomics, this study used a qualitative approach to observe and analyze the activities of a sample of evaluation practitioners. The methodology was based on two principles: comparing beginners to more experienced practitioners, and

gathering data on different practitioners working in public administrations and private firms. The sample was composed of 24 practitioners working as evaluators and evaluation managers in France and at the European Commission as well as students studying for a master's degree in evaluation. All of them were doing program evaluation as a full-time occupation (as civil servants, private consultants, or students) and claimed to be (or to be in the process of becoming) professional evaluation practitioners. They were selected on a volunteer basis and according to their experience in evaluation. To clarify the practitioners' occupations, I have distinguished between two main types of positions: (a) "evaluation managers" (in public administrations), who are in charge of defining the terms of reference and managing the whole evaluation process and (b) "evaluators" (in public administrations, research centers or private consulting firms), who are in charge of the production of the evaluation itself (data collection and evaluation report). For more details on the sample, see Table 1 below.

After completion of a task and prescription analysis of program evaluation in France and at the European Commission, I had to choose how to observe practitioners' activity. Observing practitioners throughout an entire evaluation process would have taken too long. Moreover, as this activity is partially mental, not everything would be directly accessible to observation. I therefore chose to focus on evaluation design, that is, the formulation of terms of reference in public administrations (public administration's orders addressed to evaluators, which specify the objectives of an evaluation; the scope and program concerned; the evaluative questions and criteria; the budget and time constraints, etc.) and the design of evaluation processes by private or public evaluators. As both evaluation managers and evaluators participate in evaluation design, I decided to analyze their activity with the same reference framework. Designing an evaluation appeared to be a critical phase because it implies many choices in situations that differ each time and has an impact on the success of the whole process (even if the subsequent phases—data collection and analysis, report writing, and communication—are also important).

Work Analysis Methods Used in This Study

To observe the activity of people designing evaluations, I used the following four qualitative methods and cross-compared the results to ensure their reliability:

Fourteen exploratory interviews were conducted with evaluation managers and evaluators who had different levels of experience and were working in various contexts (see Table 1). For 2 hr, they were asked the same questions about (a) their occupation and career, (b) their everyday activities, (c) their difficulties, (d) the goals they tried to achieve, (e) their success and failure criteria, (f) the guidelines and tools they used, (g) how they had learned evaluation, and (h) what they are unable to do as a practitioner.

Three in-depth interviews were conducted based on the traces of work (terms of reference and evaluation reports) with two experienced evaluation managers and one beginner with 6 months of experience. For 3 hr, they were asked to describe in concrete terms the latest evaluations they had managed, the way the evaluation unfolded, and the way it turned out. My questions were oriented by the "explication" framework to make them describe their activity in concrete terms (Vermersch, 1996).

Field observations of experienced and beginner practitioners were conducted, and focused on two activities: building a logic model and building criteria and indicators. I recorded the entire activity to analyze it.

Four clinical interviews (inspired by Piaget's work, see 1955) were conducted with two beginners and two experienced evaluation managers, who had to undertake the same case study to draft terms of reference. I developed the case study for this research. Participants were

Table 1
Sample of Practitioners Interviewed and Observed

	14 Exploratory Interviews	3 In-depth Interviews	2 Field Observations	4 Clinical Interviews After a Case Study
Evaluation managers	9 experienced: City, 5 years of exp.; Regional council, 7 years of exp.; European Commission, 5 years of exp.; European Commission, 6 years of exp.; Ministry, 5 years of exp.; Ministry, 6 years of exp.; International agency, 6 years of exp.; Ministry, 8 years of exp.; Regional council, 10 years of exp. 2 beginners: Local public agency, 6 months of exp.; Ministry, 1.5 years of exp.	2 experienced: Regional council, 10 years of exp.; Ministry, 8 years of exp. 1 beginner: Local public agency, 6 months of exp.	1 experienced: Regional council, 10 years of exp.	2 experienced: Regional council, 8 years of exp.; Ministry, 7 years of exp.
Evaluators	2 experienced: Ministry, 10 years of exp.; Private firm, 12 years of exp. 1 beginner: Research job, 1 year of exp.		1 experienced: Expert in a private firm, 15 years of exp. 1 beginner: Junior in a private firm, 4 months of exp.	1 beginner: Junior in a private firm, 8 months of exp.
Students in evaluation (Master's level)			1 group of 9 students in an evaluation class	1 student with 6 months of coursework

Note: exp. = experience.

given information about an existing program (promotion of bicycle use in a large city of France), information about the councilors' demands, the time, and the budget. They were asked to play the role of the evaluation manager of this administration and to build terms of reference to evaluate this program (to draft evaluation goals, the program's logic model, evaluative questions, a methodology, and the evaluation steering committee). I used the same protocol throughout: participants had 2 hr to study the same case, with the same guidelines and examples. I then interviewed them with the same reference framework to establish how they had solved the case. They were asked questions about (a) their difficulties, (b) their use—or lack thereof—of the guidelines and examples, (c) the way they had performed each task and the reasons for their choices, including the information they had taken into account and their goals, and (d) their level of satisfaction with the results.

The interviews and observations were transcribed in their entirety, coded and analyzed with a grid based on the guiding theoretical framework. I have identified and compared (a) the actions described and carried out, (b) the goals, (c) the results observed, (d) the contextual factors they paid attention to, (e) the reasoning orienting the choices, and (f) the stated concepts and assertions. This enabled me to reconstruct the actors' logic underlying their activity and to compare it systematically.

Strengths and Limitations of the Results

This study is based on the in-depth analysis of three kinds of data: discourses on activities, conception activities observed in real situations (actions and verbalizations during action), and conception activities observed in a simulated situation (actions and verbalizations during action and post action). The data crossing enabled me to make a systematic comparison of what these practitioners did, said they did, and said about what they did.

Although the sample was as representative of French evaluators as possible, I lacked time to study a larger sample of practitioners throughout the entire evaluation process. Not all the 24 participants were interviewed, observed, and given the same case study because of time and feasibility constraints. Even though the sample relates only to the French and European Commission contexts, I believe that the results can lead to interesting discussions with evaluation practitioners and researchers in other countries. I will now present some of these results related to the focus of this article, that is, contributing to a description and understanding of the logic of evaluators' choices in action.

Results

Different Forms of Activity

Two ways of making evaluation choices. The observations revealed that the beginners and the experienced practitioners carried out different kinds of activity. I have grouped them according to two ways of making evaluation choices when designing an evaluation. In this respect, I was inspired by Patton's terms to describe evaluators' strategies (1986). He contrasted the evaluators that he called "method-oriented" with those that he called "utility-oriented" (1986). I have changed the meaning of his terms: I have classified the less experienced as *method-oriented*, in another sense of the word (see below), and the most experienced as *results-oriented* rather than "utility-oriented," also in a different sense of the word (*Results* means evaluation conclusions and effects; it does not make any assumption as to whether the main concern is related to their quality, reliability, etc., or to their "utility").

On one hand, the beginners in this study were method-oriented because they tried to follow the prescribed goals and methods step by step, in an imitative way. They did not know when to prefer one method to another, they had difficulties in identifying and managing some dimensions of the situations in which they found themselves, and they failed to anticipate the consequences of their choices. As further results will show, they were mainly concerned by what could be called the technical matters in evaluation. Following Vincenti's results concerning engineers (1990), we could say that the evaluation beginners acted in a more mechanical and "blind" way than the experienced ones.

On the other hand, the experienced practitioners in the sample were results-oriented because they played with the prescribed goals and methods according to their diagnosis of the situation, in a more flexible way, and in relation to the final goals they were trying to achieve. Hence, they made compromises between several goals and the final results they anticipated they might obtain. They appeared to be more concerned with the larger matters or the political matters in evaluation rather than the merely technical ones.

The fact that experienced and inexperienced practitioners neither made the same situation diagnosis nor tried to achieve the same goals will be explored further. To present the results, these two dimensions (diagnosis and goals) will be discussed separately before analyzing how they are strongly linked in people's conceptualization of evaluation.

Experienced practitioners made a broader diagnosis. The level of experience in program evaluation appears to change the way people "read" (as Schwandt [2007] would say) the situations they are in. The results show that their diagnoses differed: the experienced practitioners took more note of the state of a situation; they interpreted it more and tried to transform it. In comparison, the beginners noted fewer contextual factors; they had more difficulties in interpreting and acting on some of the dimensions of the situation. While solving the same case study, the beginners mainly focused on the explicit demands—in terms of knowledge of the program—and the program's objectives, to evaluate its main impacts from a technical point of view. The more experienced practitioners also paid attention to implicit and conflicting expectations. Their aim was to anticipate the main problems they expected to encounter and to do so in a way that would ensure that their evaluation would be accepted and used by the actors, from a political point of view.

The systematic comparison of all the contextual factors taken into account by the practitioners shows that the most experienced ones made a broader diagnosis. Some contextual factors they noted were only relevant to them; others were identified by multiple evaluators. All the contextual factors they noted have been grouped into three main dimensions of an evaluation situation. They focused mainly on the characteristics of the following:

The objects to be evaluated. Level of formalization of a program's objectives (i.e., are they clear or *fuzzy*?); number of objectives and their political implications for the organization (i.e., are they *sensitive*?); number, scale, novelty, schedule, and level of achievement of a program's measures; time of the expected effects, level of knowledge about them, number of target audiences, existence of competing factors influencing the effects, accessibility of the target audiences (i.e., *silent audience*) and *easiness* to measure the changes, and so on.

People's strategies concerning evaluation. Origin of the evaluation demands (whether or not the evaluation was compulsory, whether it originated from a political and/or an administrative level, etc.); explicit and implicit demands (program renewal, accountability, monitoring, program critic, etc.); clarity, scale, compatibility of the demands; attitudes toward evaluation

(*favorable climate, habits to evaluate, hostility, etc.*); level of methodological requirements and constraints and so on.

The means to proceed with the evaluation. Scale of the resources (time and budget, existing information about the program and its effects, per diem cost, existing data collection and analysis tools on the subject, etc.); capacity and will of public actors and citizens to participate (*motivation, knowledge of evaluation processes, etc.*), and so on.

This classification builds on Patton's list of 20 "situational factors" to be taken into account in an evaluation, such as the "number of stakeholders to be dealt with" (Patton, 1986, pp. 316-317). All Patton's factors could be related to the three main dimensions of an evaluation situation presented above, for instance "staff attitude towards evaluation" could be related to people's strategies, "time available for the evaluation" to the means, and "nature of the program treatment" to the evaluated object. So these three dimensions could be considered the main dimensions of an evaluation situation, which influence practitioners' activity.

This finding shows that the difficulties to be dealt with in evaluation situations are not only technical, composed of programs to be evaluated and the means to do so (how to measure and interpret such a program, within the available resources, etc.), they are also fully social, depending on relationships with actors from different areas, with different positions, conceptions and strategies (what do they want from evaluation and do they all want the same thing? How can the actors be satisfied? How can pressure to influence the results be avoided or rejected? etc.). This finding confirms a result of a study by Fitzpatrick and colleagues (2008): "Many people think evaluation is all about methodology. These interviews illustrate the many other factors that influence evaluation" (chapter 14).

In this study, the social part of a situation appeared to be the most difficult one for beginners to interpret, whereas experienced evaluation practitioners paid close attention to it. They tended to make a broader analysis of demands, in which they not only paid attention to explicit demands but also examined implicit or conflicting expectations, different attitudes toward evaluation (more or less positive) and different methodological requirements (based on their own habits, epistemological conceptions, etc.). It appears that the experienced practitioners are also part of the situation: they take into account their own methodological requirements, capacities, and preferences.

Experienced practitioners made a more active diagnosis. The second difference stems from the fact that the experienced practitioners not only noted more situational factors, they also tried to transform the situation in a more active way. In their activity, they did not take the evaluation requirements for granted but rather questioned them. In this sense, they not only identified the stakeholders' demands but also tried to work on them: they identified them, they got people to clarify and specify them, they discussed them, and they compared different demands (see the example in Table 2 below, which was chosen for its clarity). They also worked on people's conceptions of evaluation and on their motivations for participating in the evaluation process: they described specific actions such as "making people agree," "explaining," "illustrating," "convincing," "making evaluation more attractive," and so on. These actions are directly related to the social and collective dimension of the evaluation situation under consideration.

The experienced practitioners not only adapted their activity to a given situation, they also tried to adapt the situation to their activity by transforming it. In fact, they were trying to make the objects *evaluatable* by modeling them; to create a favorable *climate* to make the evaluation process *work* and its conclusions acceptable; and to choose and create relevant means and data

Table 2
Actions Performed

An experienced evaluation manager's work on evaluation requirements at a Regional Council (France)

To define the terms of reference in her administration, the practitioner first creates a logic model of the program with the civil servants concerned. She sets out to distinguish different kinds of requirements: "When a manager comes to see us asking for an evaluation, we say 'wait a minute, we will first rebuild the program's logic to see what it is about,' to check whether we have to do an audit or an evaluation"

She uses the collective rebuilding of the program's logic to negotiate the requirements, limiting them, reformulating them, and justifying them, even sometimes refusing them: "Sometimes after having rebuilt the logic model we restrict the evaluation field because we realize, 'oh my god this is too vast, we will never be able to do an evaluation on such a huge thing, it's impossible. . . . 'My colleague and I, we then present it to the councilors to see if it also meets their expectations, to explain why we came to this, and to avoid discontentment afterwards, telling us we didn't take their opinion into account and so on, so we get into an explanation and if they say they want that, we say 'no, you will have it in another way, this is not the object' so we get on better and we won't have to face later discontentment."

She justifies the negotiation on the basis of her own evaluation expertise: "We have to deal with different demands, when we design an evaluation, we always do the splits and we never know how to deal with what is asked, what would be better and what is possible. . . . Us, evaluation managers, we have to make the councilors understand that their expectations are not always possible, that we can't do just any evaluation."

collection tools to be able to answer the evaluative questions. This work on demands and conceptions is characteristic in service activities. As studies in ergonomics have already shown (Caroli & Weill-Fassina, 2007), experienced providers of a service diagnose the client's expectations and they also discuss and negotiate those expectations. They can even redirect them drawing on their own expertise. In this sense, evaluation could be defined as a service activity implying co-activity. Part of the practitioners' expertise lies here: they have the capacity to ensure that the evaluation demands are expressed, to discuss them and even to renegotiate them from their expert point of view, arguing feasibility constraints, for instance.

The diagnosis and work on people's strategies appeared to be the most difficult for the beginners. Even though the experienced practitioners also found it difficult, many of them stated that they had learned to deal with it through experience. As an evaluation manager in a Ministry said "Beginners cannot anticipate it [the difficulties related to the actors] because they discover it when it happens. To anticipate it, you need a lot of experience. . . . Young evaluators are easily convinced of evaluation interest but it does not mean that they have the capacity to convince." Another evaluation manager in a Ministry also asserted that: "You can be manipulated by different kinds of people, make promises you cannot keep, you have to manage this and anticipate the problems. You can only do it after a lot of setbacks." Thus, experience seems to provide specific resources to deal with the social and collective part of evaluation processes. Therefore, expertise would not only be based on the capacity to use methods and tools to observe and evaluate a program but also on the capacity to manage the relations with different kinds of actors.

Experienced practitioners made more predictions. The experienced practitioners not only made a broader diagnosis of a situation, they also made more predictions about its evolution. They were trying to forecast and anticipate the evolution of a situation, and how that evolution of the situation may or may not depend on their own actions. When designing an evaluation, they made assumptions about: (a) the way people would react, (b) the probable state of the program, and (c) the way the evaluation could change direction (see Table 3). It is important

Table 3
Situational Factors Taken into Account to Make Predictions

The case study showed that the experienced practitioners made more predictions than the beginners did to solve the case. This is illustrated in the following examples.

An experienced evaluation manager building evaluative questions: he pays attention to the requirements, the logic model of the program and the monitoring information about the outcomes. He assesses the program's state: "with the diagram I can understand the program's logic and identify where the impacts are not guaranteed. And the monitoring data gives me elements on risks and difficulties . . . I presume that most of the difficulties come from the design of the routes, because we are in a domain where people have different conceptions of urban transport. I'm not sure they have really discussed this point. So I'll put the first question like this: . . ."

He also pays attention to people's strategies and makes a prediction based on their attitudes: "The [evaluative] questions will come to the councilors, if I begin to talk about management problems, to say that there are conflicts between some and others . . . my question might be refused, although it is central."

A beginner building evaluative questions and evaluation methodology: she pays attention to the logic model of the program and the formal way to draft an evaluative question according to the guidelines ("I'm on the right track.") She has difficulties anticipating the consequences of her choices: "even if it is a question, I have absolutely no idea about the way we could answer it." She also has difficulties in choosing a method: "I've got to admit, I still don't really understand which method's got to be used for what, it's still a bit fuzzy." She finally chooses the data collection tools she "knows."

to note that, from the design phase of an evaluation, practitioners made hypotheses about the program, and the way the expected impacts may or may not have been produced.

Another benefit from experience would thus be the capacity to anticipate the evolution of a given situation and to make a hypothesis on what has probably "worked" or has not worked in a certain type of program. The role of anticipation in experts' choices has already been underlined by several studies in psychology (see Chi, et al., 1988). It seems to be particularly crucial during the evaluation design phases, because the choices made then will strongly affect the whole process. This anticipation is nevertheless difficult because some dimensions of situations are dynamic: they can evolve outside practitioners' actions (like political agendas) and this dynamic nature suggests that many adjustments to the evaluation will be necessary throughout. An evaluator in a public organization recognizes that they do not control everything: "Evaluations are long processes . . . so there is always the risk of obtaining results that don't correspond to current concerns. Sometimes a topic is put on the agenda although no one expected it. Those are things that one can't control." Another evaluator comments that: often the choice of stopping or carrying on with a program isn't based on the evaluation . . . it's a more political decision. And the dissemination of reports is often limited despite the good quality of the work." This raises questions about the reliability of the predictions in relation to the relevance of cognitive resources, as we will see below. Why were the experienced practitioners who were observed making such broad diagnoses of the situations and exactly what consequences did it have on their choices?

Experienced practitioners made more compromises between several goals. The differences in the diagnoses could be explained by the fact that the experienced practitioners were oriented by broader preoccupations in their activity. They were concerned by a set of preoccupations broader than the program measure, which included it and transformed it. On one hand, the observations revealed that the beginners' goals were lower than those of the experienced practitioners and focused on technical matters of concern. They had few goals, and these were strongly connected to prescribed methods and their different phases: "to judge the effects

of the program,” “to add knowledge about it,” “to draft well-written evaluative questions,” “to build precise indicators,” and so on. They were set on adhering to prescriptions. The case study showed that the beginners often consulted guidelines (whereas the experienced evaluators never did) to find “directions,” to imitate “formulations,” and to check whether they were “doing well”: “I checked whether I was in line with the definition of the indicator, I was, so I carried on,” “I followed the orders,” “I think that for the formulation of the third evaluative question I didn’t get it right.”

The observations also show that the experienced practitioners, on the other hand, had more goals, including political considerations, which they expressed in rich and personal ways in comparison to their assessment of the situation: they tried “to judge the effects” in a way “that works,” so “it is read” and it “has effects” on the stakeholders. They also tried “to manage without setting the place on fire” in complex actor systems, “to help to sort things out” and “to put one’s finger on the problems.” As a private evaluator said, “there are often conflicts to keep things secret and we are here to help them sort it out”—all while respecting time, budgetary, and even profitability constraints.

As a consequence, the experienced practitioners made more compromises between various conflicting goals in a situation. For instance, they tried to build evaluative questions that corresponded to demands but were also “feasible.” They wanted to ensure the production of “sound” evaluation conclusions, which were “as useful as possible.” In this way, they redefined the final goals of an evaluation according to a specific situation, as already observed by Fitzpatrick (2004), purpose is “heavily influenced” by context.

Their multiple success and failure criteria can therefore be related to different sources: scientific and methodological criteria related to (a) social science disciplines and the evaluation community itself (a “solid” or even “beautiful” evaluation); (b) social criteria related specifically to the targeted evaluation audience (be “read,” “accepted” and “used”); and (c) service criteria related to the intervention logic itself (being acknowledged as a good professional, sometimes being “profitable” as in private firms, and so on).

The observation of several evaluation practitioners’ activities has thus shown that, depending on their level of experience, they had different ways of making choices related to different goals and diagnoses of situations. This has already been identified by several studies: Dreyfus and colleagues (1986) have stated in their model of skill development that people’s choices become more context dependent through experience and that experts’ thinking is “interpretative.” How could experience explain such differences in evaluative activities?

Evaluation Practitioners Have Built Different Conceptualizations of Situations

The different forms of activity can be explained by different sets of cognitive resources. These are sometimes qualified as “practical knowledge” in evaluation studies (Donaldson & Lipsey, 2008). Based on our theoretical framework, we could qualify them as conceptualizations of situations (Vergnaud, 1990), including what could be called practical or pragmatic knowledge, but also pragmatic concepts (Pastré, 2002) and action rules orienting people’s choices. As shown in ergonomics studies, the forms of conceptualization built through experience change the way people build and solve a problem, and they often remain implicit and linked in a continuous conscious flow—even if they have a strong influence on people’s activities and can be expressed under certain conditions. Vergnaud (2001), for instance, observed that even highly educated experts in rocket engineering had difficulties putting their expertise into words. This analysis has revealed the existence of particularly decisive forms of conceptualization in evaluation activities.

Table 4
Reasoning Using Pragmatic Concepts

An experienced evaluation manager at a Regional Council building evaluative questions: "If the evaluative question is too broad, we won't do it, because, we [the evaluation managers], we begin to measure the consequences in terms of data collection . . . Automatically we wonder if we have the information available, and if not, can the evaluators have it easily or not . . . so we restrict the field because we say 'the consultants, they can't make it' . . . I know what a per diem costs, and if I know that they will have to rebuild a complete data base . . . I know how much it is. So if I have a €90,000 evaluation and if the consultants will have to rebuild a data base, I will just have one question, otherwise they can't sort it out." This reasoning shows how the practitioner is preoccupied with the feasibility of the evaluative questions. She uses indications of the situation related to the means to do an evaluation (data available, time and budget, per diem cost) and the people's expectations (scale of the evaluative question asked in first intention) to anticipate what will or will not be feasible, to ensure the evaluation will be feasible.

An experienced evaluator in a private firm building indicators with a beginner to evaluate the impact of a training project: (to the beginner) "What did you write?" "Number of training sessions built with the guidelines produced by the project." "Well, it's a good indicator for itself but it's not feasible and the term 'built with' is not precise enough to be measurable . . . What would you do if you really wanted to count it? On what basis? You're not going to do a study in the whole of Europe . . . If you don't know where your information is going to come from and who is going to give it to you, there is no point in talking about indicators." This reasoning shows how the practitioner is preoccupied with the feasibility of the indicators he chooses. To make his decision, he uses contextual factors related to the object to be evaluated and the means to conduct the evaluation (a small budget evaluation).

An experienced evaluation manager at a Regional Council building a steering group (in the case study): "or I remove the councilor, because he might put pressure on the evaluation, or some actors might reject it arguing that it has once again been controlled by councilors so it does not have any value . . . Well . . . but in this case, I fear an opposite effect, councilors will not be interested and they will not take the results into account . . . Experience shows that when I have councilors in a steering group I have a better appropriation of the conclusions . . . so I keep the councilor." This reasoning shows how the practitioner is preoccupied with the legitimacy of the evaluation conclusions. He uses contextual factors related to people's strategies (conflicting demands and previous rejection of evaluation by certain elected members) and the object (the actors in charge of the program) to anticipate their reactions and make compromises between impartiality and utility concerns.

Experienced practitioners have built action rules to make their choices. The first way to explain different kinds of activity is to pay attention to people's reasoning in action. Unlike the beginners, who were trying to identify and then follow general and prescribed rules (I have to do this), the experienced practitioners often had conditional reasoning (if . . . then . . .) orienting their choices, such as "If I observe this and this and this, then the situation is like this; If the situation is like this, then I'd better do that; If I do that, then this will happen" (see examples in Table 4 below). These are experts' forms of reasoning, also called action rules when they are generalized. The experienced practitioners observed and interviewed have built their own action rules. They adapt the way they evaluate a program to a given situation (the specific state of the objects, the means to conduct the evaluation, and the political context of the evaluation, as seen above), even if they are also oriented by their preferences and habits (which are also part of the situation). Fitzpatrick (2004) has already observed that some evaluators are more able or willing to adapt their favorite evaluation approaches to stakeholders' demands, whereas others are more attached to following their own methodological preferences.

Action rules are the most visible part of practitioners' expertise. According to professional didactics (Pastré et al., 2006), one can assume that they are developed through feedback on an individual's activity and its results, in a reflexive process characteristic of learning-by-doing.

Table 5
Assertions-Examples of Pragmatic Knowledge about Situations

In regards to people's strategies, an evaluation manager in a Ministry maintained that: "you have the managers who want it to be quick, good and clear, you have the civil servants who want evaluation to help them in financing other projects, so sometimes evaluation is an alibi, and you have the influence of consultants who promise you the earth and finally give you undefined stuff."

In regards to the program's logic, while building a logic model, criteria, and indicators, an evaluation manager in a Regional Council noted that: "there are many political objectives, intentions or ambitions expressed at the same time but they will not happen at the same time, some will happen after the others . . . our work is to put into a logical order what civil servants have sold to their councilors or what they have understood about their expectations . . . between outputs and outcomes, there are also external factors, it's a bet on cause-effect links and you can have outputs that cause outcomes you didn't expect at all." These assertions directly orient the way this evaluation manager reads the program's objectives to design the logic model and chooses criteria and indicators to evaluate it (taking "external factors" into account.)

In regards to the means to conduct the evaluation, a private evaluator maintained that: "If there aren't any figures, economists don't think it's worth anything. They don't trust qualitative methods."

As Shadish and colleagues (1991) posited, it could be assumed that evaluation practitioners have drawn lessons from "their success and failures" by building action rules on the most relevant options in a given situation, to achieve a certain type of goal. Action rules can also be stated and written, so they may have been transmitted by colleagues, and/or formal and informal tutoring and guidelines. An analysis of French guidelines in evaluation shows, however, that they present a general step-by-step procedure rather than action rules telling beginners how to adapt the procedure to variations in situations. As the identification of action rules can be useful for trainers and may become a specific training objective (building the capacity to reason in particular circumstances, as we will see below), we will try to go further in analyzing what organizes them.

Experienced practitioners have built pragmatic knowledge to interpret situations. The different action rules mentioned depend on different kinds of pragmatic knowledge about the evaluation situation. The observations show that the experienced practitioners, depending on their profile and specific experience, did not interpret a situation in the same way. Their interpretations were based on what Shadish and colleagues (1991) called their own (implicit) theories. They had specific knowledge about the following (see examples in Table 5 below):

The way people usually act and react concerning an evaluation, their kinds of demands and the way they will use the results or not (related to the people's strategies).

The way different types of program usually happen and work: the way their objectives are decided and their outputs can produce outcomes (related to the objects to be evaluated).

The ways different methods can produce different results, their cost and time, their bias and relevance (related to the means to do the evaluation).

These three components of practitioners' own implicit theories can be compared with what Shadish and colleagues (1991, p. 35) have called a theory of evaluation. They identified five components of this theory: "social programming," "knowledge use," "valuing," "knowledge construction," and "evaluation practice." These results suggest another way to group these components together, according to the three main situation dimensions practitioners have to deal with in their activity (see above).

One could, however, assume that beginners had difficulties interpreting evaluation situations because they had little pragmatic knowledge about these dimensions. They had a limited framework for interpreting, making sense of and predicting the evolution of what they could observe. The systematic comparison of the interviews and activities has shown that the beginners expressed very few assertions about the situations. By contrast, the experienced practitioners expressed more personal assertions that oriented their diagnoses, predictions, and choices.

How did practitioners build such pragmatic knowledge? The existence of evaluators' own implicit theories raises questions about their origin and their degree of validity and relevance. As Demailly (2001) stated, evaluation practitioners have an "implicit sociological theory" about the course of programs and actors' behaviors. One can assume that parts of these theories have been developed through experience by observing recurrent elements in various situations, in what is called a generalization process, characteristic of learning-by-doing (Pastré et al., 2006).

Parts of them may also have been expressed and heard, becoming common beliefs shared in a given group or culture. As Lave and Wenger (1990) have shown, learning is situated and learners can be involved in a "community of practice," which embodies certain beliefs and behaviors to be acquired. In these cases, their reliability can be questioned. One can assume that they can be operative in a given context but may also be limited to the generalization of particular cases, becoming insufficient if the professional changes his or her work context. They could also be influenced by common beliefs on people's behaviors, programs, and evaluation production. For instance, an evaluation manager said that "councilors never read the evaluation reports." In this way, pragmatic conceptualizations built through experience can also be limited or even counterproductive if the persons concerned always interpret certain aspects of the situation through a rigid frame. We can suppose that stating that "councilors may not read the evaluation reports depending on . . ." would induce a more flexible and relevant activity.

Some personal theories may also stem from social science theories, seen as formal bodies of knowledge in the social sciences, such as sociology, economics, and political science. An evaluator in a public administration highlights this process: "You need to have reliable assumptions about actors' behaviors . . . If you don't have any knowledge about the way firms run nowadays, if you don't know that human resources managers usually have less importance than financial managers, you can misunderstand their talk on employment policies . . . so you know where they speak from, otherwise you can have difficulties to decipher what happens." In this way, social science theories can enrich, justify, or even contradict the implicit theories practitioners use to interpret the situations they are in. This also applies to a formal theory of evaluation. Formal knowledge can thus provide reliable assumptions on people's strategies, programs, and the course of evaluations, provided that they are integrated into people's pragmatic conceptualizations.

Hence, the formal knowledge in evaluation is not completely disconnected from the pragmatic knowledge used in action. Through experience, evaluation practitioners may have integrated parts of it into their own conceptualizations, in what is called a "pragmatization process" (Pastré, et al., 2006). We have observed that some practitioners have drawn assumptions and concepts from formal bodies of knowledge, adding pragmatic properties to them, such as "ability to explain phenomena," "relevance" according to various situations, and so on. Experience can help people to integrate formal knowledge into their own pragmatic conceptualizations, as the same evaluator notes: "when you begin, you have difficulties understanding actors' action logics, to have plausible assumptions about their behaviors . . . experience causes you to say that this assumption explains many things about what happens, the

possible effects on this program can be that and that . . ." For instance, the regular reading of evaluation reports and the lessons drawn from evaluations undertaken can help practitioners to build knowledge about the conditions for a program to achieve its objectives and an evaluation to be used.

Experienced practitioners have built pragmatic concepts related to evaluation feasibility and legitimacy. The set of pragmatic knowledge identified in this study is organized around a large number of concepts from different fields. It confirms that program evaluation is a "multidisciplinary" activity (Kishchuck & Long, 1997) in which evaluation practitioners use concepts and knowledge from various specialty fields such as (a) social sciences methods ("behavior," "external factors," "bias"); (b) public program description and analysis ("social needs," "objectives," "outcomes"); (c) evaluation process itself ("judgment," "criteria," "indicators," "benchmarks"); (d) but also consulting and intervention activities ("terms of reference," "report"). The aim of this study was not to list them all but to understand their role in people's activities.

This work analysis shows that certain concepts are more important than the others: they are used to organize people's reasoning in making compromises and predictions when designing an evaluation. In fact, the major design choices (evaluation approach, methods, questions, criteria, indicators, and data collection and analysis tools) seem to rely on the use of a limited number of concepts. They relate to two main conceptual fields.

The feasibility of an evaluation: the experienced practitioners observed tried to diagnose what would be "feasible," "measurable," "possible" or "impossible" and "evaluable" or "hardly evaluable" in a given situation; they also tried to achieve/prevent it with their choices;

The legitimacy of an evaluation process and its conclusions: the experienced practitioners also tried to diagnose and build what would be considered "sound," "valid," "reliable," "credible," or "robust" conclusions, as opposed to "not serious" and "shaky" conclusions as well as "useful," "relevant," "acceptable" and "independent" conclusions that would not be "rejected."

These concepts seem to organize the design of an evaluation according to each situation (see Table 4). The experienced practitioners used them as cognitive tools to diagnose a situation and make compromises between various goals and options. They can therefore be called pragmatic concepts (Pastré, et al., 2006) and could thus be considered the cornerstone of the activity. Even when they were specified, practitioners were not entirely aware of their prime importance in action.

The beginners in this study did not state such concepts or, the few times they did, the concepts had a limited influence on their choices. Although the experienced practitioners frequently used them, they gave different meanings to them, according to their actual position and own experience. The results show slight variations in the use of these concepts in practice, among beginners and among experimenters. Each one construed them in their own way, which tended to reduce the gap between beginners and experienced practitioners, which the analysis tended to contrast. The practitioners thus constructed the concepts' properties in different ways. The meaning of what a *reliable* method is depends, among other things, on epistemological conceptions related to disciplinary fields. Hence, the meanings of such pragmatic concepts can be developed in different ways by practitioners, relative to their levels of experience but also to their career.

In Fitzpatrick's study (2004), the US evaluators whom she interviewed were also preoccupied by legitimacy concerns. For instance, Riccio said that "using randomly assigned control groups helped us enormously in establishing our *credibility*." Rog also noted that "we wanted

to give each person the opportunity to have her voice, but we also wanted some standardization so that we could have more *reliable* information. . . [Legislators] would attach more *credibility* to quantitative data.” These evaluators had legitimacy concerns relevant to their own conception or to what they believed the stakeholders’ conceptions were (they asserted that quantitative data would provide more *reliable* information than qualitative data). Fitzpatrick and colleagues (2008) conclude as follows: “Several of these evaluators’ methodological choices were influenced by their view of what their audience expected or valued” (chapter 14). They also noted that: “Quite a few of those interviewed argued that the methodology they used needed to meet the expectations of audiences whom they believed valued *scientific* data” (chapter 14).

Why the concepts of feasibility and legitimacy appeared to be decisive in evaluation activities. Why would these pragmatic concepts be so important in evaluation activity? How is it that different practitioners, working in different positions and organizations, used similar pragmatic concepts, even if they sometimes gave different meanings to them? This could be explained by the similarities of the situations they have to face and the problems they have to solve.

Evaluation studies take place in a field of strong temporal, technical, and budgetary constraints. Program evaluators have to build knowledge and judgments about objects as complex as social change, which occur on a large scale, and are potentially hard to measure and interpret. They have to do so with limited time and money. Analyzing a situation with concepts related to feasibility would be helpful to determine the most relevant evaluation goals and methods according to these constraints.

The legitimacy of an evaluation and its conclusions is never certain but has to be created, justified, and sometimes defended. As a program evaluation is intended to influence programs’ budgeting and objectives, its conclusions may be rejected, criticized, or ignored because politicians, civil servants, and/or citizens have a strong investment in the programs’ current operations. Evaluation is strongly linked to power and actors’ strategies around public program and policy definition (Barbier, 1985). Analyzing a situation with concepts related to legitimacy would be helpful to anticipate the kind of evaluation process and conclusions that may be accepted, acknowledged, and able to produce effects. I have chosen the concept of legitimacy to cover all the practitioners’ concerns expressed above, because it offers particularly rich implications. It relates to actions of justification and their corollary and actions of acknowledgment referring to competing sources (political, administrative, scientific, professional, etc.). Evaluation legitimacy, in the first sense of the word (i.e., conclusions’ reliability or validity), is thus a function of the objects to be evaluated and the means to do so: practitioners are preoccupied by a possible bias in the program’s observation and causal analysis, and by methodological justification. Evaluation legitimacy, in the second sense of the word (conclusions’ use, impartiality, and acceptance), is a function of people’s strategies concerning evaluation and the means of the evaluation process as such: practitioners are concerned about possible pressure on the conclusions, their expected and unexpected uses, and so on.

This would explain why considerations about evaluation legitimacy, as compared with those about evaluation feasibility, have appeared to be decisive in the experienced practitioners’ reasoning. We have already noticed that they had their own, sometimes differing, definitions of these concepts. Different conceptions of evaluation legitimacy seem to be particularly decisive in practitioners’ conceptualizations and, as a consequence, in their activities. For instance, an evaluator will not make the same choices if he or she thinks that only a “reliable” (that is a “sound,” “scientific”) evaluation will help to improve programs, or that a “reliable” evaluation will not automatically be “used” and “accepted” compared with other matters of concern.

Thus, we better understand the role of evaluation theories in practice, in the second sense of the word (various prescribed methods, also grouped in “paradigms.”) They offer ready-made methods to conceptualize an evaluation situation and to act according to it. The different approaches seem to be oriented by differing conceptions about evaluation legitimacy. They are ways to prefigure and orient practitioners’ choices in the same manner in each situation. Patton has blamed paradigms for confining practitioners in limited types of activity:

The evidence from social and behavior science is that in other areas of decision-making, when faced with complex choices and multiple situations, we fall back on a set of rules and standard operating procedures that predetermine what we do, and that effectively short-circuit situational adaptability. . . . That has always been the function of scientific paradigms. Faced with a new situation, the scientist (unconsciously) turns to paradigmatic rules to find out what to do. (Patton, 1986, p. 318)

He criticizes the “comfort” of paradigms, the principles of which remain unknown and questionable. The activities of practitioners that were observed in France were complex, using ways and means related to different paradigms; in other words, they did not refer to a specific approach or evaluation theory. This may be peculiar to the French context and culture. Two of the practitioners observed had even developed pragmatic knowledge to choose either one approach or another, depending on the situation: “If the administration is really corporate (the word used by the interviewer, in English), as we say, if it is very hierarchic, it is clear that you can have an expert evaluation producing effects, if the boss is convinced, he will decide. But in local and partnership policies, experience has shown that pluralist evaluations have more effect.” So the different evaluation theories could become instruments in people’s activities. Practitioners could choose an approach according to its relevance in a given situation. In this way, paradigms and evaluation theories are also situated.

A Contribution to Evaluators’ Skills Description

By reconstructing the logic underlying evaluators’ activities, this study can contribute to the description of evaluators’ skills. However, these findings are relevant only to the design phase that has been studied in a specific context.

According to this study, designing an evaluation could be described as requiring the capacities to diagnose an evaluation problem and to build relevant evaluation designs.

The capacity to diagnose an evaluation problem. This capacity is seldom covered in evaluation guidelines or else is hidden behind technical matters (such as defining the evaluation goals, rebuilding the program’s objectives, etc.), although it appears to be a specific activity and a part of practitioners’ expertise. The capacity to identify the problem in a given situation appears to be oriented by feasibility and legitimacy concerns. It is based on the collection (through interviews, readings, etc.) and interpretation of many contextual factors including (a) the people’s strategies concerning evaluation, (b) the objects to be evaluated, and (c) the means to proceed with the evaluation. Depending on these elements, a situation will offer different levels of complexity. The evaluator has to be able to identify what will be crucial in each case: is it a problem of data collection on a particularly complex or unknown program? Is it a problem of reaching consensus on a politically controversial policy? Is it a problem of evaluation use within a tight time constraint?

The capacity to build relevant evaluation designs. This capacity is therefore strongly linked to the diagnosis of a situation. It implies the capacity to *work* on the problems (and not taking

their first formulation for granted): negotiating the demands to ensure evaluation feasibility and legitimacy, convincing the actors, etc. It also implies the capacity to make the most relevant methodological choices to treat a specific case, according to the objects to be observed, the time and budget constraints, epistemological expectations, people's will to participate, and so on.

These design capacities are not compatible with an automatic imitation and application of ready-to-use methods. Various methods can, however, be adapted to each particular case, even if the evaluators are also oriented by their own preferences and habits. Fitzpatrick and colleagues (2008) have already pointed out that: "In some cases, those choices reflect the context of the evaluation, including characteristics of the program, its culture, and the stakeholders; in others, they reflect the proclivities and preferences of the evaluator" (chapter 14).

What kind of knowledge is embodied in these skills? We have identified the existence of specific pragmatic knowledge. It confirms that evaluators have built their own implicit theories on the three main dimensions of an evaluation situation. This knowledge is different from formal knowledge but not completely disconnected from it. The two can be interwoven to enrich each other, as shown above. We have thus questioned the reliability of pragmatic knowledge and shown the interest of enriching/discussing it with formal knowledge.

To conclude, evaluators' knowledge is composite. This is consistent with the conclusion of the taxonomies of skills (see above) because the specificity of program evaluation—compared with other activities—appears to rely on the knowledge combination itself. We have identified the sources of evaluators' expertise in four different fields of knowledge, which are combined in various configurations in the evaluators' conceptualizations: (a) social science methods; (b) public program description and analysis; (c) evaluation methods themselves; and (d) consulting and intervention rules. It can therefore be assumed that a general methodological skill in evaluation would have no sense without specific knowledge about the objects to be evaluated and the social contexts in which to do so. It follows that evaluation teachers would not automatically be skilled in program evaluation, and program evaluators would not automatically be skilled in learning evaluation in classrooms.

Lessons for Evaluators' Training

This research has contributed to reflecting on evaluators' training with the firm Eureval (see Tourmen & Toulemonde, 2005). The question is no longer how to transmit evaluation methods and theories, but rather has evolved into how to develop reasoning capacities (including action rules, concepts, and knowledge) to help beginners to identify a problem in a particular case and to build relevant solutions.

The transmission of evaluation methods, theories, and guidelines seems an obvious step. However, this study suggests that it may not be enough to help beginners to develop pragmatic conceptualizations. As Flyvbjerg and Sampson (2001) have already noted, "The first rules are necessary for gaining initial experiences, but the rules quickly become a barrier to the learning process." It can confine novices in a procedural imitation (which is "comfortable," see Patton, 1986) by letting them believe that they just have to follow the guidelines to evaluate—whereas the aim is actually to help beginners to develop capacities to deal with real situations and to go beyond the transmission/imitation of ready-to-use methods. How can that be accomplished?

The role of experience in skills development has already been underlined by several studies on evaluators' training. They have emphasized the importance of experience and dialogue with the most experienced practitioners. Trevisan (2004) has listed the methods used to train young

evaluators: simulation, role play, single-course projects, and traineeship. Levin-Rozalis and Rosenstein (2003) have also presented a 1-year course based on a "learning-by-doing" approach. The aim was to "make the tacit explicit," to favor the expression of experts' "tacit knowledge" or the "conceptualization of practical knowledge." In the same way, Christie and Rose (2003) have reported the benefits of a "discussion group" of beginners and experienced evaluators at the American Evaluation Association. They have formalized action rules and noted that while beginners had learned a lot from the discussions, experienced practitioners had also enriched their experience through its explanation and discussion.

To go further, how could training sessions enhance skills development in evaluation? I will present didactic solutions that address this question.

Working on success and failure conditions. Before presenting one or several evaluation methods for beginners, it might be relevant for them to work on the final goals and the conditions of failure and success of an evaluation. For instance, they could work on case studies describing evaluation stories, then analyze the quality of the conclusions produced and their effects, and identify the conditions for success or failure in each case. The study of evaluation failures (bad quality, reports were or were not used, unused or rejected evaluations) would also be instructive. This would be a first step to help beginners to conceptualize evaluation legitimacy and feasibility.

Working on situation variations. Second, evaluation trainers usually make beginners practice a method through a case study. Trainers could make the case change so that the beginners can learn what is important to take into account and how to deal with variations in a situation. For instance, a trainer could change some dimensions of a case study: "What if you had 50% less budget? What if the program was not finished yet?" In this way, the trainers do not put too much emphasis on a canonical method but make beginners practice it in various contexts, according to various goals. In this way, the trainers can formalize action rules: "If . . . then . . .," as parts of people's expertise.

Working on social science theories. It also seems important to create links between formal teaching in social science and actual practice. Examples could show how a program and actors' strategies can be understood with different assumptions from different bodies of knowledge. Trainers could also make beginners express their own knowledge about it after a first experience, and then discuss and enrich it with formal knowledge or their own pragmatic knowledge. They could for instance ask: "What kinds of program are the hardest to evaluate? What do people usually think about evaluation? Why do some programs fail to achieve their objectives? Do interviews and case studies provide reliable data?" They could also discuss such polemic assertions as: "qualitative methods fail to provide reliable data; the objectives of a program always respond to social needs," and so on.

Working on evaluation theories. Trainers could also help beginners to make their own way through the complex evaluation literature and theories. They could ask them to solve the same case study by means of several approaches: "What would you do if you had to do this evaluation in an expert way or in a collaborative way? What if you used different types of methodology or evaluation theories? What would the most relevant and efficient approach be in this case?"

Working on people's strategies. Finally, beginners could also be trained to manage the collective part of the activity. The current study has shown that they have particular difficulties in dealing with this dimension. Trainers could discuss people's strategies around evaluations,

according to their own experience, and make the beginners aware of their complexity: “Why do people sometimes challenge evaluation results? What do they have to gain and lose from it? How could an evaluator be aware of the possible perverse effects of an evaluation process?” They could also make them develop arguments to deal with it: “What would you say to a stakeholder who rejected the conclusions? How could you present an evaluation project to civil servants who aren’t specialists?”

Conclusion

We have begun the exploration of evaluation practitioners’ pragmatic knowledge via an analysis of practices. We have focused on the reasons for their choices during the design phases. What have we learned?

First, we have restored something of the complexity of real evaluative activity. This activity consists of major and minor choices, back-and-forth movements, difficulties in decision making, compromises, contexts that are not completely under control nor easy to foresee, and so on. The observations have underlined all the dimensions the practitioners had to deal with, and all the compromises they had to make to try to “get by.” The beginners that we observed had difficulties coping with these difficulties.

Second, we have begun the exploration of the pragmatic knowledge that experienced practitioners have built through experience to try to manage this complexity. By comparing it in our sample, we have characterized its common elements and begun the exploration of its differences. This study has shown that the experienced practitioners’ pragmatic conceptualizations pertained to three main dimensions of evaluation situations: (a) the objects to be evaluated, (b) the means to conduct the evaluation, and (c) the people’s strategies around it. They have developed pragmatic knowledge on these dimensions as well as action rules. Their conceptualizations are organized by pragmatic concepts related to the feasibility and the legitimacy of an evaluation and its conclusions. These concepts have appeared to be particularly crucial in their activity. Depending on their profile and level of experience, the evaluation practitioners may give different meanings to them. This induces different kinds of strategies and actions. As Patton (1986) and Shadish and colleagues (1991) noted, there is no method better than another in the absolute, but there are more or less relevant methods depending on the context. The way evaluators will interpret how to *get by* depends on their own conceptualizations, which may be influenced by evaluation theories, but also by the lessons they have drawn from experience. Apart from the French, do evaluators throughout the world use common ways of thinking? If so, to what extent are the pragmatic concepts’ meanings (feasibility and legitimacy) influenced by evaluators’ profiles and positions?

We now better understand the role of formal knowledge in evaluation practice and in skill development. On one hand, prescriptive evaluation theories offer ready-made methods to conceptualize an evaluation situation and to act accordingly. They can guide people into a single way of evaluating or become instruments in their activity if they are adapted to situations. On the other hand, social science theories—and a descriptive evaluation theory—can offer reliable assumptions to help people to interpret the situations they face and to forecast what might happen. They can be enriched or enrich—or even contradict—the assumptions derived from the own implicit theories that evaluators have built through experience.

Finally, we have shed light on the links between the pragmatic forms of knowledge/concepts used in action and the academic and formal knowledge/concepts disseminated in the literature. The two are not disconnected, so the detour via evaluators’ practices and pragmatic conceptualizations opens wide perspectives for research and theorization in evaluation. Because

evaluation practitioners develop pragmatic theories orienting their choices, how could they be discussed and enriched by an academic theory of evaluation? However, how could an academic theory of evaluation develop on the basis of pragmatic theories? In this sense, an academic theory of evaluation would be a systematic set of descriptive assumptions about evaluation processes, actors' strategies, and public programs. As Shadish and colleagues (1991) affirmed "the ideal (never achievable) evaluation theory would describe and justify why certain evaluation practices lead to particular kinds of results across situations that evaluators confront" (p. 31). Our results suggest the need for dialogue between pragmatic and academic theories of evaluation, which could enrich each other. Taking Shadish and colleagues (1991) literally when they say that "the pragmatic concepts developed in practice probably constitute the most important basis for academic theories" (p. 35) would encourage research on the links between evaluation theory and practice.

References

- Alkin, M. C. (2004). *Evaluation roots*. Thousand Oaks, CA: Sage.
- Alkin, M. C., & Christie, C. A. (2004). An evaluation theory tree revisited. In M. C. Alkin (Ed.), *Evaluation roots*. Thousand Oaks, CA: Sage.
- Barbier, J. M. (1985). *L'évaluation en formation*. Paris: PUF.
- Bonniol, J. J. (1981). *Déterminants et mécanismes des comportements d'évaluation d'épreuves scolaires*. Thèse de Doctorat es Lettres et Sciences Humaines, Bordeaux, 1981.
- Bruner, J. S. (1990). *Acts of meaning*. Harvard: Harvard University Press.
- Caroli, S., & Weill-Fassina, A. (2007). En quoi différentes approches de l'activité collective des relations de services interrogent la pluralité des modèles de l'activité en ergonomie. *Activités*, 4, 85-98.
- Chen, T. H. (1990). *Theory-driven evaluations*. Newbury Park: Sage.
- Chi, T. H., Glaser, R., & Farr, M. J. (1988). *The nature of expertise*. Hillsdale: Laurence Erlbaum associates.
- Christie, C.A. (2003). What guides evaluation? A study of evaluation practice maps onto evaluation theory. In C. A. Christie (Ed.), *The practice-theory relationship in evaluation. New directions for evaluation Vol 97* (pp. 7-35). San Francisco: Jossey-Bass.
- Christie, C. A., & Rose, M. (2003). Learning about evaluation through dialogue: Lessons from an informal discussion group. *American Journal of Evaluation*, 24, 235-243.
- Demailly, L. (2001). *Evaluer les politiques éducatives*. Bruxelles: De Boeck.
- Donaldson, S. I., & Lipsey, M. W. (2008). Roles for theory in contemporary evaluation practice: Developing practical knowledge. In I. Shaw, J. Greene, & M. Mark (Eds.), *Handbook of evaluation*. London: Sage.
- Dreyfus, M. L., Dreyfus, S. E., & Athanasiou, T. (1986). *Mind over machine the power of human intuition and expertise in the era of the computer*. New York: The Free Press.
- English, B. (2002). Competencies for evaluation practitioners: where to from here? *Journal of the Australasian Evaluation Society*, 2, 13-15.
- Fetterman, D. M., Kaftarian, S. J., & Wandersman, A. (1996). *Empowerment evaluation. Knowledge and tools for self-assessment and accountability*. Thousand Oaks, CA: Sage publications.
- Fitzpatrick, J. L. (2004). Exemplars as case studies: Reflections on the link between theory, practice and context. *American Journal of Evaluation*, 25, 541-559.
- Fitzpatrick, J. L., Christie, T. A., & Mark, M. M. (2008). *Evaluation in action: Interviews with Expert Evaluators*. Thousand Oaks, CA: Sage.
- Flyvbjerg, B., & Sampson, B. (2001). *Making social science matter: why social inquiry fails and how it can succeed again*. Cambridge: Cambridge University Press.
- Furubo, J. E., Rist, R. C., & Sandhahl, R. (2002). *International atlas of evaluation*. New Jersey: Transaction Publishers.
- Guba, E. G., & Lincoln, Y. S. (1989). *Fourth generation evaluation*. NewburyPark: Sage Publications.
- Hadji, C. (1989). *L'évaluation, règles du jeu*. Paris: ESF.
- King, J. A., Stevahn, L., Ghore, G., & Minnema, J. (2001). Toward a taxonomy of essential evaluator competencies. *American Journal of Evaluation*, 22, 233-235.
- Kishchuck, N., & Long, B. (1997). *Rapport sur la certification professionnelle*. Canada: Rapport au bureau de la SCE.
- Lave, J., & Wenger, E. (1990). *Situated learning: Legitimate peripheral participation*. Cambridge, UK: Cambridge University Press.

- Levin-Rozalis, M., & Rosenstein, B. (2003). A mentoring approach to the one year evaluation course. *American Journal of Evaluation*, 24, 245-259.
- McGuire, M., Zorzi, R., & Perrin, B. (2002). Projet de la société canadienne d'évaluation à l'appui de la défense des intérêts et du perfectionnement professionnel, *Société Canadienne d'Évaluation*. <http://www.evaluation.francophonie.org>.
- Mertens, D. (1994). Training evaluators: unique skills and knowledge. *New Directions in Program Evaluation*, 62, 17-27.
- Nadeau, M. A. (1988). *L'évaluation de programme, théorie et pratique*. Québec: Ed. Les Presses de l'Université Laval.
- P. Pastré. (2002). L'analyse du travail en didactique professionnelle. *Revue Française de Pédagogie*, 138, 9-17.
- P. Pastré., Mayen, P., & Vergnaud, G. (2006). Note de synthèse sur la didactique professionnelle. *Revue Française de Pédagogie*, 154, 145-198.
- Patton, M. Q. (1986). *Utilization focused evaluation* (2nd ed.). California: Sage.
- Piaget, J. (1955). *The child's construction of reality*. London: Routledge and Kegan Paul.
- Rodrigues, P. (2002). Avaliação da formação pelos participantes em entrevista de investigação. Porto: Fundação Calouste Gulbenkian.
- Schön, D. A. (1983). *The reflective practitioner: How professionals think in action*. New York, NY: Basic Books.
- Schwandt, T. (2007). The role of practical knowledge in learning. Paper delivered at the 21st Annual Conference of the American Evaluation Association, "Learning to evaluate, evaluating to learn." Baltimore, 11/08/2007.
- Shadish, W. R., Cook, T. D., & Leviton, L. C. (1991). *Foundations of program evaluation theories of practice*. Newbury Park: Sage publications.
- Stufflebeam, D. (1980). *L'évaluation en éducation et la prise de décision*. Ottawa: NHP.
- Torres, R., Preskill, H. S., & Piontek, M. E. (1996). *Evaluation strategies for communicating and reporting: Enhancing learning in organisations*. Thousand Oaks, CA: Sage.
- Tourmen, C., & Figari, G. (2006). La référentialisation: une façon de modéliser l'évaluation de programme, entre théorie et pratique. Vers une comparaison des approches en France et au Québec. *Mesure et Evaluation en Education*, 29, 5-25.
- Tourmen, C., & Toulemonde, J. (2005). Comment former à évaluer? Quelques apports de la didactique professionnelle. *Education Permanente*, 165, 87-99.
- Trevisan, M. S. (2004). Practical training in evaluation: a review of the literature. *American Journal of Evaluation*, 25, 255-272.
- Vergnaud, G. (1989). Learn mathematics? In Vergnaud, G., Rogalski, J., & Artique, M (Eds.), *Proceedings of the 13th International Conference for the psychology of Mathematics Education*. Paris: International Group for the Psychology of Mathematics Education.
- Vergnaud, G. (1990). La théorie des champs conceptuels. *Recherches en didactique des mathématiques*, 10, 133-170.
- Vergnaud, G. (2001, May). Forme prédicative et forme opératoire de la connaissance. In Jean Portugais (Ed.), *Actes du Colloque GDM 2001 La notion de compétence en enseignement des mathématiques, analyse didactique des effets de son introduction sur les pratiques de la formation*, Montréal.
- Vermersch, P. (1996). *L'entretien d'explicitation en formation initiale et en formation continue*. Paris: Esf.
- Vincenti, W.G. (1990). *What engineers know and how they know it*. Baltimore: John Hopkins.
- Vygotsky, L. S. (1962). *Thought and language*. Cambridge, MA: MIT Press (Original work published 1934).
- Worthen, B. R. (1994). Is evaluation a mature profession that warrants the preparation of evaluation professionals? *New Direction for Program Evaluation*, 62, 3-15.