

A Pragmatic Guideline for Evaluation of Social Intervention

Isabel María Herrera Sánchez, José María León Rubio, and Silvia Medina Anzano
University of Seville

Considering the many theoretical and methodological viewpoints in the field of evaluation, a guideline was established to facilitate the evaluation of social intervention programs. For this purpose, the goals of the evaluation were taken into account: (a) planning interventions, (b) learning and continuous improvement of interventions, (c) programming policies, and (d) transformation of society. These goals will determine the perspective of the analysis selected (focusing on the evaluand or the context) and the strategies for change employed (focusing on processes or results). The elements that, according to Shadish, Cook, and Levinton (1991), constitute the theory of evaluation (evaluand, value, building of knowledge, and uses) have also been considered. The analysis of all these components led to the elaboration of a guideline to orient the practice of evaluation from a pragmatic perspective, in accordance with the demands and needs of a certain social context.

Keywords: program evaluation, social intervention, evaluation theory, evaluation practice

Considerando la multiplicidad de perspectivas teóricas y metodológicas existentes en el campo de la evaluación, se ha elaborado una pauta que permita conducir la evaluación de programas de intervención social. Para ello se han tenido en cuenta los propósitos u objetivos que puede perseguir la evaluación: (a) planificación de intervenciones, (b) aprendizaje y mejora continua de las intervenciones, (c) programación de políticas, y (d) transformación de la sociedad. Estos objetivos determinarán la perspectiva de análisis adoptada (centrada en el evaluando o en su contexto) y las estrategias de cambio utilizadas (orientadas a los procesos o a los resultados). Se han considerado también los elementos que para Shadish, Cook y Levinton (1991) constituyen la teoría de la evaluación (evaluando, valor, construcción del conocimiento y usos). El análisis de todos estos componentes ha permitido elaborar una guía que oriente la práctica de la evaluación desde una perspectiva pragmática acorde a las demandas y necesidades que se planteen en un contexto social determinado.

Palabras clave: evaluación de programas, intervención social, teoría de la evaluación, práctica de la evaluación

The purpose of this work is to provide a guideline to orient evaluators about the best way to perform an evaluation in the area of social intervention, depending on the goals, as these determine how the main evaluation elements are set up. However, before going into details and to make our proposal more comprehensible, we shall define social intervention as a series of planned strategies to achieve changes in the social reality in a certain population, so that this population will reach higher levels of development and well-being. Thus, the evaluation of any social intervention program involves acquiring knowledge about the value of the program and transferring such a judgment to solving social problems.

The rationale of our proposal is based on evaluators' difficulty in choosing the evaluation procedure that best matches their demands and needs, given the quantity of available strategies and their diversity (e.g., Chelimsky & Shadish, 1997; Stufflebaum, 2001).

A researcher's first task is to translate emerging demands and needs into goals. These goals can be grouped into one of the following categories: (a) planning interventions, (b) learning and continuous improvement of interventions, (c) programming policies, and (d) transformation of society. Inclusion in any category will determine the analysis outlook (micro or macro) assumed by the researcher and the change strategy adopted (linear or nonlinear).

The "micro" perspective focuses on structural (resources, personnel, etc.) and operational (activities, strategies, etc.) characteristics of the intervention program that is the object of evaluation, in order to determine which aspects are improvable. The "macro" perspective addresses the political, economic, organizational, or psychosocial aspects of a program (see Shadish, 1987).

Change strategies refer to the time period in which the evaluation takes place (at the end or during the entire process) and to the type of evaluation performed (process- or outcome-focused). According to Veney and Kaluzny (1991), linear strategies lead to punctual evaluations performed to achieve concrete goals that are usually aimed at presenting accounts and increasing the efficacy or impact of social programs. They are therefore typically performed at the end of the interventions. Nonlinear strategies consider evaluation and intervention to be part of the same process that goes from the intervention to the evaluation, and back again, so that a continued evaluation of the implicit processes of the program is performed in order to introduce the pertinent changes to optimize the intervention.

Obviously, evaluations may have various goals so that in one evaluation, several different perspectives of analysis and change strategies may be adopted. For example, an evaluation may propose to: (a) verify whether the programmed activities are being adequately implemented to achieve continuous improvement (micro perspective, nonlinear change); (b) determine the efficacy of the program so as to decide whether to continue it, modify it, or eliminate

it (micro perspective, linear change); (c) evaluate the differential effect of diverse programs with similar characteristics on different settings (macro perspective, linear change); and (d) assess the extent to which various programs that are being carried out promote democratic values such as solidarity, tolerance, and respect of human rights (macro perspective, nonlinear change).

In short, the goals of an evaluation determine the analysis perspective adopted and the way the evaluand is expected to change. However, it still remains to be seen how to perform the evaluation to achieve the proposed goals. This leads to the examination of the elements that make up the evaluation, as they will determine the decisions to be made.

For this purpose, we chose the evaluation theory formulated by Shadish, Cook, and Leviton (1991), which we think is one of the most solid and complete theoretical formulations. It compiles the contributions of the chief evaluation theoreticians of the times—Scriven, Campbell, Weiss, Wholey, Stake, Cronbach, and Rossi—considering the following elements: (a) the evaluand's characteristics; (b) the evaluation criteria, the participation level of the groups involved in the evaluation, the standpoint, responsibility and role assumed by the evaluator; (c) the evaluator's scientific assumptions and the methods used to obtain information; and (d) the use or influence of the information provided by the evaluation.

The examination of each one of these components will allow us to decide what to do and how to actually conduct the evaluation depending on the goals pursued. We have adopted a pragmatic standpoint for the analysis, as we believe that all the options are legitimate as long as they satisfy the needs encountered by the evaluator in the different settings in which the evaluation takes place.

The Evaluand's Characteristics

The systematization of the evaluation process involves considering, first, the object of the evaluation—or evaluand—, which, in our case, are the social intervention programs. We must therefore determine the main assumptions and mechanisms underlying the intervention: the components (assigned resources, intervention strategies and tactics, results), the expected causal relations among them, and the contextual factors that accompany their implementation (values, social norms, relation with other sociopolitical actions, pressure from various stakeholders, etc.).

The so-called *theory-driven evaluation* is a model that has focused most directly on knowledge of the evaluand. This movement emerged at the end of the eighties as an alternative to the experimental models or *black box evaluation* (Chen, 1990), because of its problematical application to the evaluation of social and community interventions. This theory is based on the notion that the design of an evaluation should rest on the logical model

that directs the program or intervention it is meant to evaluate. This means taking into consideration the processes involved in formulating the implementation of the intervention, as well as the results expected to be achieved by the intervention.

The appraisal of the processes chiefly involves determining whether the theory underlying the program allows one to propose the desired goals and whether the implementation of the program has been carried out according to plan. The two issues that the *normative theory* (Chen, 1990), the *implementation theory* (Weiss, 1998), and the *process theory* (Rossi, Freeman, & Lipsey, 1999) are expected to clarify are the theory's sufficiency and the program's adaptation to its conception

In the case of the evaluation of the results, the emphasis is on identifying the causal mechanisms that mediate between the components of a program and the results. The issues that concern the *causative theory* (Chen, 1990), the *program theory* (Weiss, 1998), and the *impact theory* (Rossi et al., 1999) are the efficacy, efficiency, and effectiveness of the program.

No doubt, this is the best type of evaluation to investigate the conception of a program and to gather exhaustive knowledge of the evaluand. However, it also presents a series of limitations, among them: (a) lack of precision in the elaboration of the logical models that guide the program, among other reasons, because the same program may be based on several implicit theories; or the evaluator, who may not be in the best situation to consider all the contextual factors of the program, may neglect relevant variables when conceiving the program (Bickman, 2000; Birckmayer & Weiss, 2000); (b) the models resulting from the research process performed by the evaluator are so complex that it may be impossible to translate them into a data analysis design (Weiss, 1997); and (c) knowledge of the conception of a program does not exempt researchers from using experimental designs to confirm it (Cook, 2000).

Despite all these limitations, we believe that the elaboration of logical models is usually one of the first tasks to be addressed by the evaluator. However, the problem consists of determining which elements of the program should be operationalized within the logical model and their importance in the series of activities that make up an evaluation. The higher or lesser degree of importance of the logical model will depend on the change strategy adopted:

1. Outcome-focused evaluation. This may not require a logical model, as it can study the effects of an intervention on a certain population, without the need of appraising the underlying mechanisms of the components of the program and the results achieved.
2. Process-focused evaluation. The evaluation focuses on the operationalization of the theory; the evaluator should determine whether the program begins with a theoretical analysis of the problem, whether it takes into account how the various groups involved (managers, programmers, technicians, users, etc.)

conceive the problem, or whether the theory is compatible with the system of dominant values, beliefs, and ideologies. For example, an evaluation may recommend the need to know whether a program or its components allow for the evaluation of the results (evaluation of the evaluability).

3. Process- and results-focused evaluation. An example is the proposal of Lipsey and Cordray (2000) for the evaluation of the results, which integrates an experimental or quasi-experimental design with theory-driven evaluation, providing more precise information about the causal mechanisms of the changes produced by the program. That is: (a) observed variability attributed to factors such as the implementation of the program (organization of activities developed and consistent with what was planned), the participants' characteristics, the type of contact with service providers, etc.; and (b) the cause-effect sequence between program components and their interrelations with program results.

Evaluation Criteria, Level of Participation of the Groups Involved, and the Evaluator's Role

Judgments are an omnipresent activity in any evaluation task. Judgment is doubtless implicit in evaluation, at least, of a comparative nature. This means one must explicitly express the logic used to establish the criteria that lead to the formulation of the conclusions of the evaluation. This emphasis on assessment and judgment of the evaluand is what distinguishes evaluation from research.

This led Scriven (1980) to establish objective evaluation criteria, developing a meta-theory that he denominated the *Logic of Evaluation*. According to this author, the information provided by the evaluation is related to certain judgments in four phases: (a) establishment of the criteria of merit that determine how the object of the evaluation will be labeled and judged as good, useful, of quality, etc.; (b) elaboration of standards or norms of functioning that specify levels or degrees of merit; (c) collecting data about the functioning of the object and comparing it with previously established functioning standards; and (d) synthesis or integration of the data within a judgment of merit or value.

This logic of evaluation is considered a tool that is universally applicable to the practice of evaluation (Stake & Migotsky, 1997). Arguments arise when choosing the most pertinent criteria to determine the value or merit of the evaluand, deciding the appropriateness of the standards or norms of functioning, and deciding how to measure the level of functioning. However, the biggest problems crop up when attempting to formulate the global judgment from the synthesis of the data collected. Thus, whereas for some authors, valid and objective inferences about values can be obtained from the data (Scriven, 1994), for others, this is

impossible because they consider the data to be a process of social construction and therefore, subject to interpretation by the people involved (Guba & Lincoln, 1989; Stake, 1983; Stake & Migotsky, 1997).

This leads us to the consideration of whether the evaluator should serve some pre-established absolute values, or just describe the values of certain stakeholders. In this sense, from a prescriptive approach, we propose the primacy of universal values that circumscribe the evaluation. The authors who defend this option, for example, House (1980), begin with the premise that when evaluators provide information about certain actions aimed at solving social problems (poverty, racism, violence, etc.), they are participating in a moral act that should contribute to justice in society in general.

Various difficulties arise if we decide to consider values as absolute. First, values are not limited to a theory or philosophical proposal, so it is impossible to justify one theory over another based on values. Second, in evaluation, the common practice is to have to deal with different values and interests that correspond to different groups of people involved in the actions being evaluated (Guba & Lincoln, 1989; Reboloso & Reboloso, 1998; Schwandt, 1997). In this sense, the standpoint of the majority considers it more practical to defend a *descriptive approach*: Evaluation is limited to describing or noting this plurality of interests, establishing the criteria of value on the basis of what certain stakeholders consider or perceive to be important or more just.

Precisely, the presence of various stakeholders that participate in the evaluation process is one of the main problems, because this plurality implies the frequent convergence of conflicting values. This situation forces one to make decisions about the need to consider conjointly the interests of all the people who participate in the action, or to choose the most relevant among them. There are several viewpoints on this issue. So, some evaluation models, such as the so-called *stakeholder evaluation*, defend the need to “holistically” consider the concerns and needs all the affected groups (Vedung, 1993). Conversely, others give priority to the values of certain groups, such as, for example, program and service managers (Wholey, 1983), people who lack power and who are not represented impartially in democratic decision making (House & Howe, 2000), or minority groups and those who act in their name (Mertens, 2003).

Along with the consideration of the various groups that are interested in the evaluative process as a source of information, the different levels of involvement of these groups in the evaluation task and the evaluator’s role are also particularly relevant. There are also several choices here. Thus, in the *responsive evaluation* model, although it contemplates the participation of the groups involved in the evaluation, the evaluator is responsible for and controls the evaluation (House, 2001). Conversely, in the *constructivist* approach of Guba and Lincoln (1989), both evaluators and the people involved share the responsibility in a negotiation

viewpoint of the evaluative process. Lastly, the well-known *empowerment evaluation* (Fetterman, 2001) places the stakeholders in the center of the evaluation process, looking for the self-determination of the people who participate in the program; that is, their capacity to determine their own life courses. This is a model of participation and transformation in which the evaluator assumes the role of promoter of the learning and change processes.

The inclusion of the groups involved in the evaluation process and the type of relation they have with the evaluator have reopened a debate among theoreticians about the standpoint that the evaluator should maintain during the evaluation process. It is a debate of objectivity versus commitment. Some authors state that evaluation should be neutral, in this case, the evaluators examine systematically and empirically the values that are relevant for the social policies and programs and, if they find discrepancy, they try to arrive at a consensus between the different positions of the groups (Mark, 2001). Conversely, other authors defend a standpoint that is more committed to a certain type of evaluation, that which advocates the specific interests of certain social groups, in particular, those lacking power and rights (Mertens, 2003).

In short, these proposals underscore the difficulty of assigning priority to the values that support the evaluation, of choosing and involving the stakeholders, and of determining the evaluator’s habitual role. Once again, it is noted that matching the evaluation to the goals can help solve this problem. Thus, if the aim is to determine the efficacy of a program (micro perspective, linear change), the values will already be specified because they are implicit in the intervention goals. Therefore, it is not necessary for the groups involved to participate in this aspect. Similarly, the evaluator’s role should be neutral because the results of the evaluation will depend on the causal inference being unbiased. However, if the goal is to identify and determine which program outcomes will be relevant for the various groups involved when subsequently designing the intervention (micro perspective, nonlinear change), then the values should be agreed on by these groups, which requires their maximum participation in the investigation process. The evaluator should then adopt a role of mediator or negotiator who collects information about the different needs and who attempts to reach agreements among the various criteria and priorities proposed by the groups involved.

Scientific Proposals and Methods

The key issue in this section is how to obtain the knowledge to judge the evaluand; whether such knowledge should be gathered using qualitative methodology such as that employed in idiographic disciplines, or whether it should be based on the administration of quantitative analysis procedures employed in nomothetic disciplines. In other

words, the issue is whether the object of evaluation is to understand the evaluand or to formulate general laws, translatable to what Patton (2001) calls the effective principles of programming social interventions.

Contrary to extreme viewpoints that regard both methodological approaches as opposites, we advocate a viewpoint that does not consider them dichotomic but complementary. This outlook states that the definition of the object of study and the goals of the investigation determine the limits of choice, both of the methodological approach and of the analysis techniques (see, for example, Cook, 1997). Therefore, although the epistemological bases of these two methodological traditions are different, there is no contradiction in their conjoint use in social investigation (Álvaro, 1995; Alvira, 1983; Reichardt & Cook, 1979). This perspective has produced two specific proposals in the field of evaluation:

1. The use of a specific type of methodology and its corresponding analysis techniques, which does not imply the acceptance of a whole series of meta-theoretical or epistemological proposals. For example, the *utilization-focused evaluation* (Patton, 1997), a model directed at using the evaluation outcome, so the choice of possible methodological alternatives (experimentation, quasi-experimentation, statistical designs, case studies, etc.) will depend on whether or not the data obtained meet the users' needs. For instance, if the goal of an intervention program was the academic integration of ethnical minority groups, it might be sufficient for those in charge to know the frequency of interactions between students from these groups and those from the majority group; in this case, statistical designs would be the most appropriate. However, if the managers needed to know the nature of such interactions, then observation in situ and in-depth interviews would be required.
2. The co-existence of contrasting methodological approaches and analysis techniques (causal and interpretive investigation) in the same theoretical proposal, as occurs, for example, in the *realistic evaluation* (Julnes & Mark, 1998). This evaluation model is based on *emergent realism*, which states that there is an external reality from which observable regularities may be extracted, which should be considered within the contextual dynamic of complex open systems that are observed indirectly. Therefore, the value of the different methods will depend on their capacity to verify the different aspects of the generative change mechanisms associated with the social programs and the diverse aspects of the context affected by the program. An example is the case study presented by Mark, Feller, and Button (1997), in which the impact of a new system of contracting personnel in the Administration was evaluated. For this purpose, a series of qualitative methods (observations and

interviews) were incorporated in the framework of an experimental design. These qualitative methods had several goals, among them: development and subsequent refining of the measurement instruments (surveys), validation of the results obtained in the surveys, explanation of the quantitative findings, and setting the limits of their generalization. Regarding the last point, in the interviews with the directors and the new employees, the authors clarified importance of factors such as professional category, salary, work conditions, and contract restrictions to determine whether or not it was necessary to implement the new contracting system.

We advocate the proposal of Patton (1997) and we coincide with his opinion that the methodology should be selected according to the users' needs, which are reflected in the evaluation goals. To avoid subjecting the evaluation to private interests, these goals should be established as a result of what the users need to know and the evaluator considers most pertinent, given the peculiarities of the object of the evaluation. Thus, at the start of the program, when little is known about the problem and its solutions, it is most appropriate to discover and describe the peculiarities of the problem (for example, by case studies). When a program is sufficiently implemented at a local level to be able to offer an account of its efficacy, the evaluation will require a causal or probabilistic explanation (experimental, quasi-experimental, or correlational designs). If the differential effects of similar programs in different settings should be determined so as to ensure the continuation of some political strategy, the evaluation should lead to generalizable conclusions (such as the case of the synthesis of research or meta-analysis). Lastly, in order to go beyond the impact of the program and take into account the value of various possible alternatives to solve the social problems, the evaluation should be of an interpretive nature, using a hermeneutic model.

In short, we believe that evaluation should allow one to establish a series of general principles applicable to any social intervention, without having to give up understanding the specific evaluand in which the evaluation originated.

The Utilization or Influence of the Information Provided by the Evaluation

Research in this field has focused on identifying how the findings of an evaluation affect the different stakeholders. Initially, the implicit use theories stated that the results of an evaluation should become the central inputs in political decision making (*instrumental use*). However, experience soon showed the evaluators that political decisions were not made according to the results of the evaluation; sometimes because the results were contrary to political interests and others because they were not definite.

Although the results do not lead to new actions or to the modification of the evaluated actions, they may be used for such ends as: (a) to review the intervention foundation, indicating possible directions for action, which Leviton and Hughes (1981) called *conceptual use*, and Cronbach (1982) and Weiss and Bucuvalas (1981) *enlightenment*; and (b) to stir up a political debate about some social problem; for example, the AIDS prevention programs promoted by the local administrations were not very effective for this purpose, but they were good for setting off a social debate about young people's new ways of experiencing sexuality. This has been called a *symbolic* or *political* use (Leviton & Hughes, 1981).

None of these notions was concerned with identifying the utilization within a conception based on the very *process* of evaluation, which is the central axis of the *developmental evaluation* proposed by Patton (1994). This evaluation can be considered a process of constant dialogue between evaluators and the parties involved, who share the responsibility to generate, transmit, and use the information. Everything produces learning in individuals during the evaluation process, that is, a process of constant feedback and change, where the evaluator works with a team of persons who want to learn how to improve what they are doing and to learn from their actions.

Patton (1997) was also interested in the intentional nature of the knowledge generated by the evaluation, so his proposal includes the identification and differentiation of the *primary users* from the *audience* in general, as these elements may prefer different usages of the evaluation. For example, a politician, as primary user, may be interested in the results of the evaluation to justify his stance regarding some social problem, whereas the citizens may be interested in the use of the results of the evaluation to improve the policy of distribution of goods.

All the above indicates the multidirectional nature of the usage of the results of evaluation, which has led Kirkhart (2000) to propose that it may be preferable to talk about *influence* instead of *use*, so as to underline the unintentional and noninstrumental effects of the evaluation. For this author, it is essential to determine the *sources of influence* or such elements of the evaluation that comprise the active agents of change and that may be associated both with the processes and the results of the evaluation, and also the *temporal dimension*, or moment in which the influence occurs, because its effects will vary depending on whether it occurs during or at the end of the evaluation, or at some interval after the completion of the evaluation.

In short, it is obvious that research of the usage has taken a new direction in the last years, going from being a minor issue to becoming a fundamental aspect in research. This is not surprising if one takes into account that evaluative work should meet specific requirements both of the political and the scientific context, requirements which are currently seen as interwoven, thus shattering the rational outlook of the

utilization of the results of social experimentation, which had been the predominant viewpoint since the decade of the sixties (Campbell, 1988).

An Integrative Approach

After reviewing the components of the evaluation, we propose an organizing guideline in which the various ways of performing an evaluation can be integrated. In our approach, the different combinations of these elements are implicit in the goals of the evaluation, and these combinations generate different ways to address the dimensions of evaluation, as shown in Table 1.

According to this guideline, the evaluator could adopt one or several of the following combinations, depending on the goals of the evaluation:

1. Planning interventions.
2. Learning and continuous improvement.
3. Programming policies.
4. Transformation of society.

Planning Interventions

Evaluations within the framework of this goal are meant to obtain information about the results achieved by the development of a program, an activity, etc. An example of this type of evaluations is the *planned variation evaluation*, in which different varieties of a program are compared to determine their differential efficacy (see, for example, Yeh, 2000).

The goals of these evaluations define a micro perspective and a linear change strategy; that is, the evaluation focuses on the evaluand, with the aim of obtaining information about the efficacy, efficiency, or effectiveness on which to base a decision about future planning. This means eliminating actions that have proven to be inefficient or implementing actions that have been shown to be efficient.

This type of evaluations, supported by methodological rigor as a quality criterion, requires the objectivity and independence of the evaluator, who is responsible for the entire evaluation process. When accepting the standards of objectivity and independence, the information provided by the evaluation is expected to be relevant for the selection and adoption of the best solutions, although other unintentional influences may co-exist. In this case, in addition to improving the efficacy of a program, the evaluation may be useful for other latent purposes, such as increasing the visibility of the program in the eyes of the community.

Learning and Continuous Improvement

The purpose of this type of evaluations is learning and continuous improvement of the intervention. These evaluations focus on the analysis of the correspondence between what

Table 1
Guideline for the Evaluation of Social Intervention Programs

Dimensions	Goal of the Evaluation			
	Planning interventions (micro perspective, linear change)	Learning and continuous improvement (micro perspective, nonlinear change)	Programming policies (macro perspective, linear change)	Transformation of society (macro perspective, nonlinear change)
Evaluand	Programs, interventions or actions performed	Development of the program, action, or intervention	Series of interventions with common goals	Actions that lead to discrimination and inequality
Evaluation criteria	Efficacy, efficiency or effectiveness	Quality and implementation	Impact	Social justice and inclusion
Level of participation	The audience is informed or consulted	Participation in part of or in the entire process	The audience is informed or consulted	Active participation.
Evaluator's standpoint	Independence, objectivity	Examines the values	Independence, objectivity	Defense of values
Evaluator's responsibility	Maximum	Shared with the involved groups	Maximum	Minimal
Evaluator's role	Scientific	Moderator	Resolute	Promoter
Scientific assumptions	Standards that account for causality	Emphasis on the specification of the context	Searching for generalizable conclusions regarding the influence of the context	Hermeneutical cycle of interpretation
Methods or considerations for the design of the evaluation	Experimental, quasi-experimental, or correlational	Emerging designs, mixed designs, case studies	Meta-analysis. Synthesis. Good practice	Dialogue. Critical self-reflection
Influence of the evaluation	Focused on evaluation results. Promoted at the end of the evaluation.	Focused on evaluation process. Promoted from the start.	Focused on evaluation results. Promoted at the end of the evaluation.	Comprises the final goal of the evaluation

was intended and what was actually implemented so as to carry out the necessary adjustments to produce the desired change in the goal of the evaluation. Therefore, they adopt a micro perspective and a nonlinear strategy.

An example of this type of evaluations is the above-mentioned *theory-driven evaluation*, which consists of the formulation of a theory that describes the evaluand's structural and operational characteristics in detail, in order to provide information about its nature, functioning, and aspects that may be subject to change. For a review of the most recent applications of this model, see Donaldson (2003).

Contrary to the previous perspective, this type of evaluation has a formative nature and the participation of those responsible for the implementation and initiation of the program being evaluated is essential. They should play a relevant role in the specification of the evaluation issues, collecting information, interpreting the results, etc. This

participation has obvious effects on the support for the program and on its evaluation, directly affecting decision making and problem solving.

The stance that the evaluator should adopt about the values of the groups involved can vary from distant to near. In the first case, the evaluator just compiles, describes, and examines these values neutrally. In the second case, the evaluator tries to understand the meaning of the program for the persons involved, because this meaning reveals which aspects of the program should be changed to improve its functioning.

Independently of the evaluator's standpoint, he or she works in collaboration with the groups involved, moderating and clarifying the discussion, facilitating decisions on the basis of the data obtained.

Lastly, from this perspective, the methodological choices will depend on the nature of the intervention and on the degree of specification and formalization of the intervention.

Thus, for example, when problems are not well defined, people disagree about which aspects should have priority, the goals are diffuse, or the activities are not implemented in their totality, then the evaluator should start out with naturalistic assumptions so as to discover and understand the phenomena being studied in their context. According to this viewpoint, it is not necessary to begin with a prior design, but instead a design “emerges” and takes shape as the evaluation advances; quantitative measures can subsequently be incorporated to increase the credibility of the findings (see for example Kalafat & Illback, 1998). On the contrary, when the program is highly specific and formalized, one should choose quantitative methods from the start.

Programming Policies

In this group are the evaluations whose purpose is to formulate general intervention principles to help improve social policies and compile the best practices or lessons learned. In this case, the evaluations go beyond the limits of a concrete program or action, adopting a macro perspective aimed at political, social, organizational, and psychological factors that have facilitated or hindered the achievement of the best results (linear strategy).

There are two options here: the first consists of performing a meta-analysis of the individual works and integrating the findings obtained by each of the particular evaluations and revealing effects that would otherwise not be observable (see, for example, Lipsey, 1997; Lipsey & Wilson, 1993). The second option consists of comparing a series of interventions with common aims and characteristics but taking place in different contexts. This type of evaluation is known as *cluster evaluation* (see Sanders, 1997).

The purpose of both types of studies is to obtain generalizable findings that lead to better comprehension of the intervention programs by analyzing the diverse circumstances and contexts in which these interventions achieve the best results. From this viewpoint, the evaluators do not limit their role to that of mere advisors in the decision-making process to develop new initiatives, policies, or implementation strategies, but instead they also make suggestions to solve the problems.

Transformation of Society

As in the previous case, this type of evaluations are within the socio-political context, but conversely, they express misgivings about the main aspects of the programs in order to promote substantial changes in values and political priorities (macro perspective). Thus, they generate a permanent debate about the processes that reveal the weaknesses of the current social system (nonlinear change strategy) so as to solve the most urgent social problems: poverty, isolation, delinquency, etc. For example, the main

purpose of the *feminist evaluation* is to make explicit the elements that explain gender inequality (Sielbeck-Bowen, Brisolaro, Seigar, Tischler, & Whitmore, 2002).

For this purpose, the evaluator adopts an activist stance that defends the interests of the groups involved, in particular of those who lack power and rights, and promotes the active participation of these groups to transform the existing structures. Conversely, the evaluation is based on debate, dialogue, and reflection within the hermeneutical cycle of observations and interpretations that lead to the understanding and comprehension of reality.

Lastly, given the critical examination of the values and assumptions of the programs and the need to make the current inequalities explicit, the evaluation plays an active role in political activity, because the goal is for the evaluation to be linked to social transformation.

Summing up, the purpose of this guideline is for the evaluation to be comprehensive and realistic; that is, to accept that the evaluation meets the demands and needs of a certain social context and the goals of solving social problems, using the evaluation theory that is most appropriate for one's purposes. This perspective, rather than breaking up the field of evaluation, should, to a great extent, contribute to the productive progress and plurality of interests, perspectives, and values of society.

References

- Álvarez, J.L. (1995). *Psicología social: perspectivas teóricas y metodológicas*. Madrid: Siglo XXI.
- Alvira, F. (1983). Perspectiva cuantitativa-cualitativa en la metodología sociológica. *Revista española de investigaciones sociológicas*, 22, 53-75.
- Bickman, L. (2000). Summing up program theory. *New Directions for Evaluation*, 87, 103-111.
- Birckmayer, J.D., & Weiss, C.H. (2000). Theory-based evaluation in practice. *Evaluation review*, 24, 407-431.
- Campbell, D.T. (1988). *Methodology and epistemology for social science. Selected papers*. Chicago: University of Chicago.
- Chelismky, E., & Shadish, W.R. (Eds.) (1997). *Evaluation for the 21st century. A handbook*. Thousand Oaks, CA: Sage.
- Chen, H.T. (1990). *Theory-driven evaluations*. Newbury Park, CA: Sage.
- Cook, T.D. (1997). Lessons learned in evaluation over the past 25 years. In E. Chelismky & W. Shadish (Eds.), *Evaluation for the 21st century: A handbook* (pp. 30-52). Thousand Oaks, CA: Sage.
- Cook, T.D. (2000). The false choice between theory-based evaluation and experimentation. *New Directions for Evaluation*, 87, 27-34.
- Cronbach, L.J. (1982). *Designing evaluations of educational and social programs*. San Francisco: Jossey-Bass.
- Donaldson, S.I. (2003). Theory-driven program evaluation in the new millennium. In S.I. Donaldson & E.M. Scriven (Eds.),

- Evaluating social programs and problems: Visions for the new millennium* (pp. 109-141). Mahwah, NJ: Erlbaum.
- Fetterman, D.M. (2001). *Foundations of empowerment evaluation*. Thousand Oaks, CA: Sage.
- Guba, E.G., & Lincoln, Y.S. (1989). *The fourth-generation evaluation*. Newbury Park, CA: Sage.
- House, E.R. (1980). *Evaluating with validity*. Beverly Hills, CA: Sage.
- House, E.R. (2001). Responsive evaluations (and its influence on deliberative democratic evaluation). *New Directions for Evaluation*, 92, 23-30.
- House, E.R., & Howe, K.R. (2000). Deliberative democratic evaluation. *New Directions for Evaluation*, 85, 3-12.
- Julnes, G., & Mark, M.M. (1998). Evaluation as sensemaking: Knowledge construction in a realist world. *New Directions for Evaluation*, 78, 33-52.
- Kalafat, J., & Illback, R. (1998). A qualitative evaluation of school-based family resource and youth service centers. *American Journal of Community Psychology*, 26, 573-604.
- Kirkhart, K.E. (2000). Reconceptualizing evaluation use: An integrated theory of influence. *New Directions for Evaluation*, 88, 5-23.
- Leviton, L.C., & Hughes, E.F. (1981). Research on the utilization of evaluations: A review and synthesis. *Evaluation Review*, 5, 525-548.
- Lipsey, M.W. (1997). What can you build with thousands of bricks? Musings on the accumulation of knowledge in program evaluation. *New Directions for Evaluation*, 76, 7-23.
- Lipsey, M.W., & Cordray, D.S. (2000). Evaluation methods for social intervention. *Annual review of Psychology*, 51, 345-375.
- Lipsey, M.W., & Wilson, D.B. (1993). The efficacy of psychological, educational and behavioral treatment: Confirmation from meta-analysis. *American Psychologist*, 48, 1181-1209.
- Mark, M.M. (2001). Evaluation's future: Furor, futile or fertile? *American Journal of Evaluation*, 22, 457-479.
- Mark, M.M., Feller, I., & Button, S. (1997). Integrating qualitative methods in a predominantly quantitative evaluation. A case study and some reflections. *New Directions for Evaluation*, 74, 47-59.
- Mertens, D.M. (2003). The inclusive view of evaluation: Visions for the new millennium. In S. I. Donaldson & E. M. Scriven (Eds.), *Evaluating social programs and problems: Visions for the new millennium* (pp. 91-107). Mahwah, NJ: Erlbaum.
- Patton, M. Q. (1994). Developmental evaluation. *Evaluation Practice*, 15, 311-319.
- Patton, M.Q. (1997). *Utilization-focused evaluation: The new century text* (3rd Ed.). Thousand Oaks, CA: Sage.
- Patton, M.Q. (2001). Evaluation, knowledge management, best practices, and high quality lessons learned. *American Journal of Evaluation*, 22, 329- 336.
- Reboloso, E., & Reboloso, J.R. (1998). Significado y desarrollo actual de la evaluación de programas. In E. Reboloso (Ed.), *Evaluación de programas: ámbitos de aplicación* (pp. 9-29). Barcelona: Textos Universitarios "Saint Jordi."
- Reichardt, C.S., & Cook, T.D. (1979). Beyond qualitative versus quantitative methods. In T.D. Cook & C.S. Reichardt (Eds.), *Qualitative and quantitative methods in evaluation research* (pp. 7-32). Beverly Hills, CA: Sage.
- Rossi, P.H., Freeman, H.E., & Lipsey, M.W. (1999). *Evaluation. A systematic approach* (6th ed.). Thousand Oaks, CA: Sage.
- Sanders, J.R. (1997). Cluster Evaluation. In E. Chelimsky & W. Shadish (Eds.), *Evaluation for the 21st century: A handbook* (pp. 396-404). Thousand Oaks, CA: Sage.
- Schwandt, T.A. (1997). The landscape of values in evaluation: Charted terrain and unexplored territory. *New Directions for Evaluation*, 76, 25-37.
- Scriven, M. (1980). *The logic of evaluation*. Inverness, CA: Edgepress.
- Scriven, M. (1994). The final synthesis. *Evaluation Practice*, 15, 367-382.
- Shadish, W.R. (1987). Program micro- and macrotheories: A guide for social change. *New Directions for Program Evaluation*, 33, 93-109.
- Shadish, W.R., Cook, T.D., & Leviton, L.C. (1991). *Foundations of program evaluation: Theories of practice*. Newbury Park, CA: Sage.
- Sielbeck-Bowen, K.A., Brisolaro, S., Seigar, D., Tischler, C., & Whitmore, E. (2002). Exploring feminist evaluation: The ground from which we rise. *New Directions for Evaluation*, 96, 3-8.
- Stake, R.E. (1983). Program evaluation, particularly responsive evaluation. In G.F. Madaus, M.S. Scriven, & D.L. Stufflebeam (Eds.), *Evaluation models: Viewpoints on educational and human services evaluation* (pp. 287-310). Boston: Kluwer-Nijhoff.
- Stake, R.E., & Migotsky, C. (1997). The evolving syntheses of program value. *Evaluation Practice*, 18, 89-103.
- Stufflebeam, D.L. (2001). Evaluation models. *New Directions for Evaluation*, 89, 7-98.
- Vedung, E. (1993). Modelos de evaluación. *Revista de servicios sociales y política social*, 30, 39-68.
- Veney, J.E., & Kaluzny, A.D. (1991). *Evaluation and decision making for health services program* (2nd Ed.). Ann Arbor, MI: Health Administration Press.
- Weiss, C.H. (1997). How can theory-based evaluation make a greater headway? *Evaluation Review*, 21, 501-524.
- Weiss, C.H. (1998). *Evaluation* (2nd Ed.). Englewood Cliffs, NJ: Prentice Hall.
- Weiss, C.H., & Bucuvalas, M.J. (1981). Truth test and utility tests: Decision-makers frame of reference for social science research. *Evaluation Studies Review Annual*, 6, 695-706.
- Wholey, J.S. (1983). *Evaluation and effective public management*. Boston: Little Brown.
- Yeh, S.S. (2000). Building the knowledge base for improving educational and social programs through planned variation evaluations. *American Journal of Evaluation*, 21, 27-40.

Received: May 26, 2004

Revision received: November 11, 2004

Accepted: November 30, 2004