

EVALUATION AND RESEARCH: DIFFERENCES AND SIMILARITIES

Miri Levin-Rozalis
Department of Education
Ben-Gurion University of the Negev
Beer-Sheva, Israel

Abstract: This article discusses the similarities and dissimilarities between research and evaluation, which are two clearly differentiated disciplines despite their similarity in concepts, tools, and methods. The purpose of research is to enlarge the body of scientific knowledge; the purpose of evaluation is to provide useful feedback to program managers and entrepreneurs. In this article I examine the central characteristics of research and evaluation (validity, generalization, theory and hypotheses, relevance, and causality) and the different roles those characteristics play in each. I discuss the different functions of evaluation and research, and propose some criteria for fulfilling the different demands of evaluation and research. And I argue that the constant pressure to examine evaluations by the criteria of research prevents evaluation from becoming an independent discipline and delays the development of standards and criteria that are useful to evaluators.

Résumé: Le présent article discute des similarités et des différences entre la recherche et l'évaluation — deux disciplines clairement distinctes malgré la similarité de leurs concepts, outils et méthodes. La recherche a pour but d'enrichir le bloc de connaissances scientifiques, et l'évaluation a pour but de fournir une rétroaction utile aux gestionnaires de programmes et aux entrepreneurs. Dans cet article, j'examine les principales caractéristiques (validité, généralisation, théorie et hypothèses, pertinence et causalité) et les différents rôles qu'elles jouent dans l'évaluation et la recherche. Je discute des différentes fonctions de l'évaluation et de la recherche, et je propose certains critères pour répondre aux différentes exigences de l'évaluation et la recherche. Je soutiens également que la contrainte constante d'examiner les évaluations à la lumière des critères de la recherche empêche l'évaluation de devenir une discipline indépendante et ralentit l'élaboration de normes et de critères utiles pour les évaluateurs.

Corresponding author: Miri Levin-Rozalis, Department of Education,
Ben-Gurion University of the Negev, P.O.Box 653, Beer-Sheva, Israel, 84105;
<Rozalis@bgumail.bgu.ac.il>

THE PROBLEM

There seems to be an ongoing question about the connections and differences between evaluation and research. After last year's meetings of the European Evaluation Society and American Evaluation Association, I decided to take up the challenge and try to clarify some of the distinctions between these two activities. This article is strongly influenced by my view of evaluation as a service to interveners and a tool for improving the running of a project.

In my fieldwork as an evaluator, I frequently encounter a lack of awareness about the essence of evaluation, in general, and the difference between evaluation and research, in particular. In the literature we often find "evaluation-research" used as a kind of hybrid term, but in many cases this is a product of miscegenation that is neither good research nor proper evaluation.

In this article I will discuss the similarities of and the differences between evaluation and research. I propose to show that these are separate and distinct disciplines despite the similarities that arise from their sharing concepts, instruments, and methods in some cases. I claim that the difficulty in distinguishing between evaluation and research is at the expense of evaluation. The ongoing attempts to apply research criteria to evaluation put pressure on evaluators to relinquish the special attributes of evaluation; when this is done, the uniqueness and quality of evaluation is diminished.

The essence of scientific research can be found in ongoing philosophical arguments — the rationalists versus the empiricists, the positivists versus those espousing symbolic-interaction — that tend to emphasize specific aspects of the endeavour (Bechler, 1987; Caws, 1965; Copi, 1961; Hempel, 1965; Kincheloe & McLaren, 2000; Kuhn, 1962; Popper, 1959). However, several criteria and definitions are accepted by all of them. It is generally agreed that the aim of research, as a scientific method, is to expose and understand the basic laws according to which the world operates, and to increase our body of knowledge in order to attain a better command of nature and to satisfy the basic human urge to acquire knowledge. Hence, for research, the knowledge that is amassed is the purpose of the activity — pure knowledge has a value in and of itself.

There are often questions raised about the definitions of "applied sciences," "research and development," and "qualitative research."

It seems that the borders between research and evaluation are not so clear in these areas. These categories are discussed below, but at this stage, I want to say two things. First, there are, of course, many points where research and evaluation overlap; otherwise, there would be no confusion and, indeed, where this occurs, the similarities are greater than the differences. Second, in all these kinds of research, any solution achieved through research and development or conclusion derived from qualitative research arises from generalization and a quest to define general laws. The differences between those and pure research are a question of extent.

As we have seen with scientific research, the essence of evaluation has not been adequately defined. However, by definition, evaluation is in a process of constant change and development because it is used for inspection and study. It is generally seen as a tool in the hands of program executives, decision-makers, or interveners. Evaluation is used in many fields, from educational and social processes to health services and agriculture. It can be used strictly in quantitative surveys or it can be totally open-ended. The function of evaluation is also very broad. It begins with the aim of evaluating the level of success of a program, determining if a given program has advanced or reached its goals and what the side effects of its activity may be, in order to improve the efficiency and contribute to the continuation of an ongoing educational, community, or social project¹ (see Cronbach, 1963, 1983; House, 1980; Joint Committee on Standards for Educational Evaluation, 1981; Nevo, 1989; Patton, 1981; Scriven, 1967; Stufflebeam, 1971; Tyler, 1950). Or it can be a “practical, material, and political undertaking concerned with examining and enhancing the ways we make interpretive judgments of the value of human actions that unfold in specific social and cultural contexts” (Schwandt, 2002, p. 2).

Interest in evaluation arose in the United States because research was not sufficient to meet the demand for a systematic examination of what was going on in the field. This demand appeared together with, and as an outcome of, the development of a systematic and structured process of decision-making that has gradually appeared in the areas of both formal and informal education and social sciences. Previously, before the 1960s, many activities in the social or educational fields stemmed from a purely philanthropic approach (Renzulli, 1975). Since then, new systems and methods of looking at projects, which differ from those of conventional research, have been used, breaking from the conventional stream of research (Rossi &

Freeman, 1982). These systems and methods were not only new in their approach but were also called by a different name: "evaluation" (Rossi & Wright, 1984; Tyler, 1942).

The demands on the evaluators and evaluation in general changed from examining operational and measurable aims in the 1950s to producing useful information for the decision-makers and even to shaping the actual intervention in the 1970s (Nevo, 1989; Scriven, 1967; Stufflebeam et al., 1974). In the same period, it became clear that, in order to respond successfully to the demands of different projects, evaluators needed to be inventive (Patton, 1981). In the 1980s, evaluators were expected to take into account anyone who might be affected by the activity of evaluation and to use the evaluation to restructure and reframe the concepts and world-views for its clients (interveners, decision-makers, and the programs' clients) (Abma, 1997; Drewett, 1997; Guba & Lincoln, 1989; Pawson, 1996), or to influence the power structure by empowering those being evaluated (Brisolara, 1998; Cousins & Whitmore, 1998; Fetterman, 1996).

These changes and demands conflicted with clear, structured research settings. The evaluator was expected to determine what kind of knowledge was important, to whom it was important, how to collect it, how to process and interpret it, to decide what conclusions had been reached, and to give advice based on those conclusions. Because of all these problems, evaluators often find themselves between the hammer and the anvil. On the one hand, they are required to provide answers to many varied questions, but on the other, they lack the clear frameworks and suitable tools needed to do this, while facing a constantly changing environment (Eyken, Goulden, & Crossley, 1995; Guba & Lincoln, 1989; Tucker & Dempsey, 1991). Some evaluators have found solutions for this problem by borrowing methods from other disciplines (see Finne, Levin, & Nilssen, 1995; Kazi, 1996; Tucker & Dempsey, 1991), others, by relating to theory (Chen, 1990; Chen & Rossi, 1981, 1992; Pawson & Tilley, 1997).

Despite the differences between evaluation and research, and despite the different demands made on the evaluator and the researcher, there is still confusion about their roles, for several reasons: it is not clear what research itself is, and, especially in the social sciences, the definition of research changes all the time. The same is true for evaluation. Often working in obscure situations, evaluators themselves are looking for a clear-cut, well-recognized frame of reference in which to work. Research procedures and criteria pro-

vide these well-recognized frames of reference, and, in addition, research is held in greater esteem than evaluation.

I summarize my claims in table 1, which I will refer to in a broader way in the next pages.

Table 1
Similarities and Differences in Evaluation and Research

	Evaluation	Research
Area of application	Application of the examination as wide as possible	Application of the examination as wide as possible
	Narrow application of findings focused in the project	Application of findings as wide as possible
	Aim of providing concrete feedback	Aim of increasing the body of scientific knowledge
Theory	Field-dependent: theory used to enlarge the understanding of findings	Theory-dependent: derived from or aspiring to theory
Methodology	Evaluation setting and data collection methods derived from the field	Research setting and data collection methods derived from theory
	The evaluator is reactive	The researcher is active
Generalization	Attempt to understand what is happening in a specific project	Attempt to formulate a general law; external validity is important
Relevance	Useful for the project	Increase of scientific knowledge
Causality	Stresses internal validity; that which is an artefact in research is seen as an internal variable in order to reach causality	Internal validity is important; stress is on a small number of causal variables in isolation from other variables

DOMAIN OF APPLICATION

The main difference between evaluation and research is the domain of application. Both evaluation and research aspire to increase our understanding in a variety of areas, but the kinds of understanding they contribute to are different. The dominant paradigm in research is meant to formulate general knowledge as general laws; thus, the

understanding it aspires to reach is that of an abstract and generalized law that can be applied to as many events as possible. (Today, in social sciences, there are various challenges to this paradigm such as the interpretative approach, the relativistic approach, post structuralism, and critical theory. But these approaches also aim to achieve some general understanding of the phenomenon at scope.) In contrast, evaluation is intended to amass knowledge and understanding of a concrete activity — a project — and to give this understanding back to the project as feedback. Evaluation attempts to investigate the mutual influences between a maximum number of variables at a given time and place.

Where research is intended to enlarge a body of knowledge (and the knowledge has value in and of itself), in evaluation, knowledge is only a means to an end and its value is first and foremost in the feedback for the evaluated project. In that sense, the evaluation process is circular: it is a cycle that begins and ends with the project. In social or educational interventions or changes in the community or in politics, follow-ups by evaluators have become a necessity. The ability to stand aside and observe, to see and to understand processes, to organize them, examine them, and return them as intelligible feedback is a serious contribution, one that improves the ability of a project or activity to be effective under changing circumstances.

When we speak of the application of evaluation or research, it is important to distinguish between two main points: the procedures of examination (questionnaires, interviews, observations) that the evaluators or researchers use for collecting data; and the findings and conclusions derived from the data, and their consequences. The reason for making this distinction lies in the principal differences between evaluation and research — in the possible application of the two methods.

When we speak of procedures, wide application is important in both research and evaluation. For the evaluator, it is therefore advisable for the examination procedures to be systematic and repeatable, able to be followed up and criticized. If statistical procedures are used, it is important to use accepted statistical methods. If possible (and in many evaluation cases it isn't), it is better to use familiar tools, or tools that can be used for other settings.

When we speak of findings, there is another difference between evaluation and research. In all kinds of research, we aspire to the

widest possible application of the findings, as well as the conclusions that can be drawn from them, so that they can serve as a base for general laws. That is the primary task of research.

In evaluation, however, the findings have to be relevant to a specific project and correct at a specific time for the project evaluated. There is no attempt to come up with findings that can be generalized beyond that project. That is not to say that, as Pawson (1996) put it, “tales” and “stories” from the field, without any conceptualization or any cycles of learning (Patton, 1997), can serve as proper findings — they can’t, because they leave the project members with the same understanding of their situation as before the evaluation came in. That is not to say that it is impossible to obtain findings or, more often, conclusions that have value above and beyond the specific project. This happens frequently, but it is not the primary task of evaluation. I come to this question of generalization later on in this paper.

In cases where evaluators use research procedures and carry out evaluations in research settings, they use research tools that are valid for focused questions. As a result, the application of the findings is broader. But there is a paradox: defined, general, and abstract variables may increase the applicability of the findings, but they decrease the usefulness of the conclusion for the given project, meaning that while the ability to generalize separate findings and variables increases, the ability of the evaluator to give intelligible, coherent feedback concerning the big picture of the project is very much decreased. What happens is that the research apparatus gives highly generalizable, abstract answers appropriate to research: valid answers gathered with the aid of highly replicable research tools. But the quality of the evaluation is damaged because the operators of the project do not get answers relevant to their own work — answers that are directly related to the different activities, audiences, and questions of the project. Highly generalized findings, composed of theoretical variables and written in scientific language, are hard to translate into operative proposals for improvement. That is why they have little or no use as feedback. (Such a report is sometime needed to justify the existence or continuation of a project, or for raising money, but not for the ongoing daily work of the project.)

The same problem exists in research. The broad application of the findings frequently interferes with a deeper understanding of the subject being researched (Geertz, 1973; Goetz & LeCompte, 1984;

LeCompte & Preissle, 1993; Sabar, 1990; Tucker & Dempsey, 1991). Research methods designed to overcome this problem, like case studies, ethnographies, and other methods of qualitative research (Creswell, 1998; Sabar, 1990), are constantly being developed. In these approaches, researchers forego the broad, wide application of their findings in an attempt to get a better understanding of the complexities of their subject. In this instance, research is similar to the approaches used to evaluate a project. They are not, however, identical, in that the goals and functions of qualitative or ethnographic research or the case study method are not changed. The research still aspires to enlarge the body of scientific knowledge by formulating generalized laws, with the broadest application of findings.

THE PLACE OF THEORY, HYPOTHESES, AND METHODOLOGY

Turner (1986) and Wallace (1971) wrote about scientific research as a means of manipulating knowledge, controlled by accepted methods and anchored in theory. Theory provides structure, generalization, definition, and propositions that represent a systematic view or representation of a phenomenon by specifying the relationships between variables in order to explain the phenomenon — or even to make predictions in similar cases (Kerlinger, 1972). In scientific research, theory is the main frame of reference of the accepted logic of research: deductive logic and inductive logic (Copi, 1961; LeCompte & Preissle, 1993).

There are writers, such as Ray Pawson and Nick Tilley (1997), who claim that this is also true for evaluation. From their very epistemological perspective, they see (realistic) evaluation as beginning “with a theory of causal explanation.” They are not the only ones. On the other side of the coin is Michael Scriven (1991), who claims that there is no use for theory in evaluation.

The “theory driven evaluation” was first introduced and later developed by Chen and Rossi (1981). They wanted a solution to the problem of experimental designs of evaluation that allows us to claim the program a success, but say nothing about why it works. According to Chen (1990), theory is the means to explain causality. It specifies the underlying causal mechanisms that link, mediate, or condition the causal relationship between the treatment variable(s) and outcome variable(s) in a program. Chen uses social science theory to support information collected from different stakeholders. As Owen and Rogers (1999) claim: “this means gathering what is al-

ready known about the social or educational phenomenon under review” (p. 195).

Pawson and Tilley (1997) claim that without a theory of why an intervention may be effective and a theory of the conditions that promote this potential, research into its usage is blind (p. 82). Following Wallace’s (1971) wheel of science, they reframe theory in terms of propositions about how mechanisms are fired in context to produce outcomes. The mechanisms, the context, and the outcomes are those of the evaluated program. They declare themselves to be pluralists when it comes to the choice of methods, but those methods have “to be carefully tailored to the exact form of hypotheses developed earlier” (p. 85).

In the next section I hope to convince the reader that causal theory is the result of an evaluation process and not its generative principle, and that evaluation methods stem from the project and not from the theory.

There are three paths of research logic that connect theory and data: deductive logic, inductive logic, and abductive logic. In deductive logic, a valid logical connection exists between the hypotheses and a previous theoretical assumption. The hypothesis is an *explanandum*, meaning it is explained by deductive premises derived from a theory. There is nothing new in the hypothesis, nor is anything new permitted. The a priori theoretical assumptions are the *explanans*, which explain the hypothesis. No matter what else may be true in the world, or what other information may be discovered, the validity of the connection between the *explanans* (a priori premises) and *explanandum* (hypothesis) is not affected. This method of formulating hypotheses is good for research that examines a theory or tries to refute it. It assures the researcher that there will be no deviation from the application of the theory in question. According to this, phenomena that appear in the field are not subject to deductive logic at all; the field is merely the court in which the a priori hypotheses can be examined (Copi, 1961; Copi & Burgess-Jackson, 1995).

Deductive logic is the opposite of the logic used for evaluation, because evaluation examines the field (the evaluatee) in order to reveal the variables and the elements that play a role and the connections between them. It does not use the field to validate variables and suppositions stemming from an existing theory. Pawson and Tilley are the representatives of such a deductive logic. Owens and Rogers

(1999) present an example borrowed from Weiss (1996): If contraceptive counselling is associated with a reduction in pregnancy, the cause of the change might seem to be the counselling. But what in the process caused the change? The knowledge provided? It might be that the existence of the counselling helps to overcome cultural taboos against family planning; it might give women confidence and assertiveness in sexual relationships; it might trigger a shift in the power relationships between men and women. These or any of several other cognitive/affective/social responses could be the mechanism leading to the desired outcome. Using deductive logic we will never know, because the theory leads us to, in Pawson and Tilley's (1997) words, the mechanisms, contexts, and outcomes that are part of its frame of reference. It is not important when examining a theory, but it is crucial when we want to know what it is that works in a project.

In inductive logic, hypotheses are formed according to generalized phenomena that are observed in the field. In an attempt to formulate a general law of probability, these hypotheses examine the probability that these phenomena will be repeated. In order to do this, we must know the characteristics being investigated in the group we are focusing on and the a priori conditions (for example, that a coin has two sides and that when it is tossed it will land on one of them) (Copi, 1961).

These conditions do not usually occur in an evaluation. First of all, empirical generalization is a claim for the reasonable appearance of a phenomenon. In order to claim empirical generalization, the characteristics of the phenomenon have to be known ahead of time — they have to have been examined earlier. Evaluation begins, at least in part, early in this examination or perhaps even before it. It attempts to lay bare the significant phenomena and thus cannot yet examine their characteristics and the probability of their occurrence. Second, the aim of empirical examination is to generalize beyond the individual case, to attempt to formulate a general law. For evaluation, for all the reasons mentioned above, this is not the main goal.

Evaluation is not theory-dependent; it is field-dependent. There are evaluation theories, evaluation models, and various approaches to questions of evaluation. There are theories that we use to interpret the findings of evaluation, but the essential element in evaluation is dependent on what we get from the field, in the sense that the questions that a particular evaluation poses do not arise from a

theory. They come from the project being evaluated, its targets, its environment and clients, the needs of the people running the intervention, the decision-makers, interested parties at various levels, and the needs of the evaluators.

The work of the evaluator is, in this sense, similar to the work of a doctor or a psychologist, who examines a particular patient, looking at that patient's unique history and conditions. Doctors and psychologists then find validation for their findings and assumptions in the available research and pertinent theories. In the same manner, evaluators base their work on wide professional and theoretical knowledge in the various fields they deal with. When the doctor is treating a patient, she is not validating a theory unless she is doing medical research, yet she relies on the knowledge gained from research. Thus, evaluators, relying on theoretical knowledge, do not attempt to validate a theory. The attempt to draw hypotheses from some specific theoretical framework will limit the scope of the evaluation and prevent the evaluator from making hypotheses that do not arise from this theoretical framework. The theoretical framework dictates the concepts we use and their expected relationships with each other. In an evaluation, a theory that is suitable to the project can be chosen at a later stage, when the evaluator is drawing conclusions and explaining the findings.

Evaluation is very much guided by the third type of research logic: abduction (Peirce, 1960). The principles of abduction are based on the notion that there are no a priori hypotheses, no presuppositions, no theorizing in advance.² Each event is scrutinized and its importance examined (Shank & Cunningham, 1996). Hypotheses are then formed about the event: is it connected to other events and, if so, how? Perhaps it is an isolated event, and if so, what is its meaning?

The explanations we form for these new events are "hypotheses on probation." A cyclical process of checking and rechecking against our observations takes place, widening and modifying the explanation through this process (Levin-Rozalis, 2000).

According to this principle, evaluators must free themselves from research hypotheses, generalized concepts, or theories that would focus their observations. This does not mean that their actions are purely intuitive or without previous knowledge. There are, of course, aims and concepts behind the intervention, and evaluators are knowledgeable and influenced by their professional background. But they

act, insofar as possible, without any theoretical bias about what they are looking for and what they will exclude. Should they look at clients' cognitive performance or social behaviour, or maybe their emotional coping? Or none of these things at all? An a priori theory tells evaluators what to look at; abduction logic tells them to look at all of these phenomena. In a certain sense this is like the work of a detective who has to be free of presuppositions and open to all the possibilities and information that the investigated subject offers. The instructions to the evaluator, as to the detective, must be the following: "Never theorize in advance of your facts ... Look at the whole scene inside and out; ask yourself what you saw, not what you expected to see or what you hoped to see, but what you saw" (James, 1989, pp. 34, 53).

I do not intend to claim that the evaluator examines every fact or event in the field. Such an examination is not possible, nor is there any need for it. I am saying only that in scientific research, theory serves as a framework that dictates which data are relevant and which are not. In evaluation, the needs of the project and the reality in which it operates determine the questions and the variables. It is advisable for the people running the project, together with the evaluator, to define the problems. But even if this does not occur, evaluators look at the situation in the field, and that situation, with all its complexity, generates the questions that they will explore.

After identifying a surprising fact in a project, it might be important and interesting to examine it in controlled research methods. In that way we can create new knowledge not only in the project context but also to contribute to the scientific body of knowledge.

The statistical variance between variables (which interfere with the work of the researcher, unless we are dealing with some sort of qualitative research) are part and parcel of the evaluation. Evaluators do not try to cull variables from the situation. Indeed, variables and phenomena that are irrelevant for checking the probability of a general law are often very relevant to the evaluation. Evaluators deal with many phenomena (which are not always variables³): interpersonal relationships, states of mind, styles of management, values, catchwords, styles of dress, and so on. We are speaking of phenomena that are specific to a project, such as the complex of characteristics of the different people involved in the project, the size or shape of the building in which the project takes place, or the neighbourhood in which the activity is carried out (without comparison with a representative sample of similar phenomena). Hypotheses are

checked systematically by means of a multi-operational approach — that is, using a variety of reactive and nonreactive measuring tools generally employed in the social sciences and the humanities: interviews, observations, questionnaires, document analyses, and so on. Only when the diverse findings point to the same conclusion can it be called an important finding.

For example, in a project that operates a number of programs, a problem arose with a lack of initiative among the paraprofessionals who worked in the programs. Several possibilities for this were postulated: the function of the paraprofessionals was not clearly defined, so it was not reasonable to expect initiative. The workers were not suited to their function; the instructions that they received were harsh and inflexible; deviation from the plan resulted in “punishment.” Centralized management discouraged initiative. It would be hard to change things in the project, a fact that created a kind of learned helplessness. The problem was related to a cultural pattern in the population who were the subjects of the program (and from whom the workers were chosen). The workers had insufficient training and were therefore uncertain and afraid to improvise. All these possibilities are hypotheses on probation.

When the field was examined, some of these suppositions were shown to be wrong. It turned out that the workers got weekly training, both individually and in groups. The working atmosphere was supportive and there was no “punishment” for overstepping the boundaries. The management were democratic and gave the workers a lot of autonomy. As these suppositions were disproved, others grew stronger, mainly the suppositions that the source of the problem lay in a cultural behaviour pattern that perceived some kinds of initiative as blunt, rude behaviour that challenged supervisors as not being good enough. From this point, we could go on and examine this partially validated supposition systematically, to see if the problem existed in other elements of the program and to search for solutions that would be appropriate for the source of the problem — solutions that could be anchored in the workers’ culture.

Evaluators lay reality bare with all the scientific means they have, in the most systematic way that can be applied to a specific project. The conditions in the field being examined dictate the instruments that can be employed, as well as the language and the character of the instrument — for example, a questionnaire is not suitable for a population that has difficulty reading. The conditions also dictate

the concepts of the examination — for example, using the concepts and language of the project rather than those of the evaluator's scientific jargon. In this sense, the activity of evaluation is like qualitative phenomenological research, but as mentioned above, it has different aims. As I stated earlier, the main aim of qualitative and phenomenological research (as with any research) is to increase our understanding and knowledge concerning humanity in general, and thus it must find variables that can be generalized and formulated into rules. In order to do this, the researcher must sieve all the variables connected to a specific event that are not subject to generalization or do not belong to the system of suggested explanations. In contrast, the activity of evaluation tries to use all the phenomena that have been discovered in the examination in order to present a coherent system of explanation for the project being examined, thus diminishing the ability to generalize the findings.

In its holistic approach, evaluation is similar to research in anthropology, but anthropological research is conducted in an attempt to reach generalizations about human behaviour and to add to the body of human knowledge (Ashkenazi, 1986). When conducted in a natural environment, it deals with variables inherent in anthropological concepts: institutions and social organizations, family structures, kinship, social activities, ceremonies, symbols, and cultural artifacts.

Evaluators organize their findings in significant and, if possible, comparable patterns. While it is often possible to generalize specific findings into more generalized structures and processes, or to reach an explanatory theory, the main goal is to find a coherent explanation for what is going on in the project, with its successes and failures.

For example, we found that the manager of a project we were evaluating had an authoritarian management style that suppressed any initiative his workers might take. Knowing this, it became possible to come up with a generalized explanation, supported by theories about the connection between authoritarianism and initiative. We could not postulate in advance that the manager was authoritarian, since we had no idea what kind of a manager he was or what kind of interactions we would find with the workers in the project, nor even if the question of management-worker relations would be pertinent. Evaluators themselves have to have enough knowledge about styles of management and about authoritarianism and initiative to develop a generalized explanation. If they don't, they can get help from a professional who does know about management styles.

The questions that direct the evaluation activity are not derived from a theory, but in many cases, the project itself is the application of a theory. The programs and methods of the project are based on the rationale of some theory, or from the *Weltanschauung* of its initiators, its operators, and its financiers. The quality of the rationale that dictates the project's activities, its connection to what happens in the field, and the question of whether the results are congruent with the rationale (or guiding theory) are among the things that the evaluation examines. In this case, the questions that the evaluators ask are in dialogue with the theory.

The methodology of classic research stems from a theory and its hypotheses and empirical generalizations. It is the fruit of pedantic attention to planning a strategy for operationalizing the hypotheses, that is to say, formulating the variables in observational terms. An observational term has to be anchored in reality; a suitable situation for observing must be found or artificially created before the appropriate scientific observations are carried out to test the hypothesis. To do this, the researcher must choose the variables that seem most suitable to the concepts of the hypothesis: the most suitable subjects, the most felicitous field and the most appropriate set-up.

Methods of evaluation are different from research methodologies. Where researchers can choose their own subject for research and propose the research approach (which we will call an active method), the evaluator is reactive and at times pro-active. Evaluators react to events and, at times, tries to direct them. Evaluators cannot choose the participants, set-ups, or variables of the project. The field is given, the participants are given, and, at least in part, the variables are not known in advance. There is a general definition of the evaluation questions, but they are not defined in terms of a hypothesis and the variables are not operational. At times, even the nominal definitions of the concepts of a project are a question for evaluation. The instruments of evaluation (interviews, discussions, observations, questionnaires, videos, analysis of protocols, dialogue analysis, or any other tool) are planned and chosen according to the population involved, the activities to be checked, the question to be evaluated, the time and money available, and the contract signed between the operators of the project or the initiators and the evaluation team.⁴

Evaluators do not examine isolated variables; they examine events, which include most of the possible variables together with their interconnections and contexts, as well as factors that are not vari-

ables (the type of neighbourhood or the character of the school, for example). In their analysis of the events, evaluators define categories and variables, based on previous knowledge and according to either the aims of the project or the finding. For example, if a variable like opacity in definition of roles occurs repeatedly in different aspects and locales of the project, it then becomes possible to see if its occurrence and influence are identical in each case, what its origins are, and so on.

Evaluators need to have a high degree of professionalism to enable them to be flexible in collecting data as well as constructing the research tools that are appropriate for the target population, the project context, and the examined issues. As in anthropological or qualitative research, the evaluator him/herself is an instrument. In addition to their academic and professional qualifications, it is important for evaluators to have the ability to make contact and communicate, for example, to win trust and to present their findings in a way that will make them not only understandable but also acceptable for application. Despite the importance of the evaluator's personality in this process, evaluation must be systematic, structured, and professional, and not "art" or intuition-based.

One question that arises frequently is why there is such stress on the actual project, and whether all projects are not, in the end, essentially the same. Patton (1981) describes the differences between projects. He lists 20 variables, including the type of project, the availability of data, the structure of the decision-making process, the educational level of the staff, and the educational level of the clients. If each of these variables has only three values (the highest, middle, and lowest), we get 8,000 possible combinations. Even if we exclude the less probable combinations, we still have a very high number. The field in which the evaluator acts keeps changing because of the differences between projects and the changes in a project itself over a period of time. It is almost impossible for a complete explanation, taking into account all the activities in the project on its multiple levels, to be the same for two separate projects.

THE VALIDITY OF EVALUATION FINDINGS

If each project is different from every other project and demands a different approach, the question of the validity of the findings and the possibility of generalizing them arises. In terms of the validity of a research setting or an evaluation setting, it is customary to say

that both research and the researcher stress internal and external validity (Cook & Campbell, 1976, 1979). Evaluators stress only internal validity since they are interested in the here and now of a specific project.

The method of evaluation overcomes, with relative ease, most of the threats to its internal validity that come from the variables being examined (e.g., history, maturation, selection, mortality, statistical interaction with selection, resentful demoralization of control group subjects, or local history). All these threats are included in the scope of the evaluation. In contrast to research, evaluation deals with a multiplicity of variables and not with isolated ones. Variables that in research would be artifacts are of interest to the evaluator.

The question of validity in relation to measurement evaluation is related to the fact that measurement in the social sciences is, with very few exceptions, indirect, because most of the concepts it deals with are abstract or theoretical constructs. This means that researchers in the social sciences can never be completely confident that what they are measuring is indeed the thing they want to measure (Nachemias & Nachemias, 1982). However, uncertainty about the construct validity of the findings (confounding or fuzziness of the theoretical variable, or lack of congruence between a theoretical variable and an observational variable) does not pose a problem for evaluation since the operational definition and the questions used in evaluation stem from the field and not from a theory. In evaluation, the construct validity of a measuring instrument means the congruence of the observational definition and the nominal definition of the variable being studied (Peres & Yatsiv, 1994).

For the past two decades, writers have tended toward a fuzzy definition of validity, permitting subjective judgements. Thus, for example, Cronbach (1983) claims that validity is more subjective than objective. Krathwohl (1982) compares validity to persuasion, reliability, and consensus. Campbell, the father of the concepts of internal and external validity, has retreated from these concepts almost completely (Campbell, 1986). He now suggests formulating new definitions. He proposes that internal validity be changed to local, molar (pragmatic, a-theoretic), causal validity. This innovation questions whether a specific bundle of treatments creates a real change in an activity at a specific place and time. According to Campbell, there are cause-and-effect relationships that appear in the field, the source of which is not explained in any theory, nor do

we have any idea how to generalize them. If we are to do so, we must reverse the usual order of putting theory before the activity. The cause and effect that are seen in the field, and which do not stem from a theory, can provide a lever to develop a new theory.

External validity touches on the question of the generalization of the findings, and on the stable connections that can be predicted between empirical phenomena. Since research is usually aimed at generalizing findings to other systems, populations, situations, and times in order to be able to formulate a general law, a typical research question would then be, "Is the development of social relations among people living together for 24 hours a day over an extended period of time applicable to different populations in other situations?" In an evaluation, however, we can say that what interests us is the consolidation of a group of candidates in a course for fighter pilots, from which the command echelon can draw conclusions. Are the same factors of interest to people who run summer camps for youths and who want to raise the level of satisfaction of the participants? In looking at fighter pilots, evaluators are not really interested in the question of whether the same connections are formed between the inhabitants of a retirement home or a boarding school, but researchers are. Thus, while the question of external validity is extremely important for research, it is not particularly important in evaluation.

On the other hand, the question of the ability to generalize the effect of a project is important. In a project, the managers or the entrepreneurs activate a program in order to produce a desired effect. Evaluation can check whether the effects of the project can be generalized or what the scope of the project's application may be, if this question is of interest. In other words, we are dealing with the possibility of generalizing the treatment activated by the project. Can it be transferred to a different locality or to a different population? We can generalize our conclusions but not our findings.

Sometimes, when a generalization is made from a sample of participants in a project to the whole group (or groups) of participants, there is a question of whether the sample was truly representative. In evaluation there is an interest in knowing whether a generalization from one part of a project can be applied to another part, from one set of variables to another set, in order to avoid describing the project as an unconnected series of events. For example, a lack of clear boundaries characterizes the definition of the tasks of senior staff in a project. Is this also a characteristic of the tasks of the rest

of the project staff? Or, there is a lack of familiarity with the definition of central concepts in a project involving a single school (i.e., a problem with a specific school). Is this really a problem of the specific school, or may that lack of familiarity also be found in the other schools in the project (i.e., has there been a failure on the part of the operators of the project to transfer their message)?

These hypotheses about the possibility of generalizing findings help the evaluator to draw a coherent picture, to discriminate between unusual phenomena (inefficiency in a usually efficient situation or vice versa, for example), and to identify the source of the problem (the operators' lack of awareness of the importance of conceptualization or their inability to stand up to teachers whom they no longer trust). The possibility of generalizing a specific finding provides evaluators with an important tool for understanding the field in which they are operating. Proof of such a possibility gives evaluators important and relevant knowledge both for themselves and for the project they serve.

An examination of a number of significant variables in a structured research setting can be a part of a wider evaluation setting that includes other examinations, additional variables, and questions about the connections between events and the effect of their mutual interactions. The operative concepts of "isolated variables," "controlled samples," and "generalizability" do not exist in the field in which evaluation operates. In practice, variables are not isolated, the group that is examined is not pure or controlled, and the possibility of generalizability is either very limited or does not exist.

A project involves various factors that affect its functioning: the operators of the project are connected to entrepreneurs, financiers, executives, other operators at various stages, and, of course, the target population, which is not without its own subdivisions. Such a varied group means there are different world-views involved, different aims, and a variety of values. There is an activated program, which can be simple and one-dimensional (such as teaching Hebrew to new immigrants) or complex (like changing socialization patterns in a community of Ethiopian immigrants). Even the building from which the project operates can have an influence — what is its distance from the neighbourhood where the target population lives? Is it open or enclosed? Warm and inviting, or cold and forbidding? All these factors and their mutual interactions are only a subset of the factors that influence a project's effect.

Most of the procedures in a research study do not relate to these factors and their interactions, as they are outside the area of application. From the point of view of research, they merely interfere. As Kuhn (1962) expresses it, these are not the sort of puzzles that well up from standard research procedures. Research procedures such as quasi-experimental designs (Cook & Campbell, 1976), qualitative research, anthropological methods, and case studies try to cope with this problem of multiple variables. While a qualitative researcher may be interested in the mutual interactions between generalized and abstract variables, or even the interactions between specific variables in a given project, the main aim is still to understand the phenomenon in order to contribute to knowledge, for the sake of the knowledge.

The influence of all of this on the success or failure of a given project is usually not the sort of question researchers ask themselves. For evaluators, however, these questions are the essence of the matter. Researchers choose their research questions according to their area of knowledge and the question's interest and importance to the researcher. Evaluators choose them according to the probable usefulness of the answers in the project they are serving: in other words, according to their relevance to the project.

THE IMPORTANCE OF RELEVANCE

The relevance of an evaluation is extremely important to its quality, since an essential part of an evaluation's value is whether it provides efficient feedback to all levels of decision-makers. The evaluator gives a service, and the service must be effective: the effectiveness of the feedback and the value embodied in it have to be relevant. Effective and relevant feedback that deals with a concrete product of a concrete project and a given group of people at a given time depends on the quality of the examination. In research, because the aim is to be able to generalize the findings, there is no question about the relevance of the findings to the field in which the research was performed (unless we are dealing with some sort of "research and development" department of a specific enterprise that has to provide answers to specific questions).

Furthermore, the questions of which subjects are relevant and which are irrelevant or where it is worth investing evaluation resources and where it is not may be a subject for negotiation between the evaluator and the project directors or the funders of the research.

Interest, in and of itself, is enough reason for research, but it is not enough reason for formulating the questions of an evaluation. In the same way, research questions are not a matter for negotiation. They stand between the researcher and her/his colleagues. The decision of how to examine the subject is reserved solely to the initiator(s) of the research.

Activating a procedure of monitored examination that deals with generalized and abstract variables does not properly serve evaluators and their clients because it does not stand up to the criterion of relevance. The answers that can be produced are far from the clients' immediate world and concerns.

In order for feedback to be relevant and useful, it must fulfil the following conditions:

Immediacy. Feedback should be immediate, or given at least within a time span that will permit the study of findings and conclusions that contribute to the creation of or planning for change. If it delayed for too long, the feedback will not be useful to the people it was intended for. Even for a summative post facto evaluation, the feedback has to be available for future planning.

Particularity. The feedback must be particular so that it can give answers dealing directly with the project under discussion. That means it has to deal with the specific audiences of the project, the operators, the clients, and the planners of the project. It should not deal with generalized concepts.

An example of this is an empowerment project involving Bedouin women. The women's frustration can be made much clearer to the people running the project if we can refer to concrete situation factors. In other words, we should not say, "The patriarchal structure of this society is threatened and the men are resisting the project because of the separate treatment the women are receiving. Something needs to be done to reduce this perceived threat and the resulting resistance." It is better to present the fact that the men are not cooperative and don't allow their wives to participate in some of the activities and that it might be better to form a men's group and explain the aims of the program to them.

Another example is from a program designed to teach contraception to women of Ethiopian origin in Israel. Feedback showing that coun-

sellors of Ethiopian origin are more successful at this than the non-Ethiopian counsellors because they speak the women's language and understand their cultural codes is better than saying, "Agents of change from within the target population are more efficient at bringing about changes in the habits of intimacy."

Results that are delivered as general laws or as abstract, theoretical concepts may be useless to the operators or decision-makers in the field. Furthermore, they are not always understood by people who are not scientists.

The language of the project and its terms. Feedback should be phrased in the terminology of the project, or at least in language that is understood by its operators and initiators, even if their definitions and terminology are not precisely congruent with the professional jargon. If the project terminology uses the term "counsellors," we will not call them "group leaders" or "change agents" or "coordinators" or "facilitators," even if we feel this is the correct term. If a project claims that it creates "empowerment" or "autonomy" in its target population, these are the concepts that we have to deal with. We can, of course, negotiate a change in inadequate concepts or try to coin new terms during the process of setting up the evaluation, but whenever specific terminology exists, we have to use it. This is true for all kinds of evaluations: formative and summative evaluations, evaluations of process and evaluations of product.

Value judgements. We cannot speak of the relevance of feedback without touching on one of the main differences between evaluation and research, that related to the relevance of findings: the judgemental aspect of evaluation. The history of evaluation shows that the essence of the development of this discipline is the evaluation client's need for the evaluators to interpret and judge the findings. This is also the origin of the name — evaluation. At the end of the 1970s, a committee was convened in the US to define the essence of evaluation as a discipline. This committee decided to stress the importance of the evaluator's decision concerning the worth or the value of the object being evaluated (Joint Committee on Standards for Educational Evaluation, 1981). This is so even in summative evaluations that focus only on the product (measuring the achievement of the aims of the project). We are frequently asked to judge whether measuring the achievement of project goals is significant (Mor, 1992). This is particularly true of the formative approaches so popular nowadays, which open to discussion the value of goals, the value of

the project, the damage or losses it is likely to cause, and the significance of its effect.

Research tries to be objective and the findings themselves are the message. Because evaluation is a tool for decision-makers, it cannot and should not avoid recommendations, even if they deal with values. The process of evaluation that involves only gathering data, without judging its essence, without advice, without at least delineating alternatives and helping to choose between them, is neither possible nor desirable (Cordray, 1986; Joint Committee on Standards for Educational Evaluation, 1981; Levin-Rozalis, 1987, 1998; Lincoln & Guba, 1986; Scriven, 1967, 1983; Stake, 1969).

THE ISSUE OF CAUSALITY

A significant element in feedback from evaluation is the attempt to understand what causes what, what stems from what, what is cause, and what is effect. Only if we know the cause can we try to control the effect, but answers about the connections between variables are not usually satisfying. Nachemias and Nachemias (1982) claim that there is a basic dilemma between causality and generalization — that in order to assure the clarity of proof of causality, we frequently sacrifice the ability to generalize. That, of course, is true for research. In evaluation, as discussed above, the ability to generalize may not be so important; however, the need to understand the cause of phenomena certainly is.

For example, children who were participating in a project (Levin-Rozalis & Bar-On, 1994) complained bitterly about the food. They called it “teachers’ food” and said that they wanted “home food.” Repeated attempts to improve the food, or to prepare more traditional food, did not help. There was no correlation between specific kinds of food and the number of complaints. This connection alone was of scant use to the people running the project. An in-depth investigation showed that the children didn’t really think that the food was that bad; rather, the food served as a channel for their feelings of dissatisfaction because other channels were blocked. Because it was considered an honour to be part of the project, the children did not dare to express their dissatisfaction. By understanding the nature of the children’s trouble and the cause of their complaints, the operators of the project were able to bring the problem into the open and deal with it. When the problem was solved, the complaints about the food stopped.

Mor (1992) argues that even summative evaluation, which is done after a project is finished, cannot be satisfied merely with results, but must examine what was done and how, why the findings are what they are, and what caused them to be so. This is not to say, of course, that the explanation of cause is always available, or easy to discover, or that the evaluators always try very hard to uncover it.

SUMMARY

Evaluation is not recognized as a separate discipline, and in many cases, people see it as bad or incomplete research that does not stand up to rigorous research criteria. This creates a vicious cycle: evaluation is not recognized as scientific research or as an academic method in its own right, which puts pressure on evaluators to try to meet the criteria of research. They try to reach higher validation rates than the paradigm of evaluation makes possible, to get generalizable findings, and to carry out investigations using already validated tools. This limits the evaluation settings and causes evaluators to examine isolated effects and abstract concepts, resulting in findings that are adequate neither as feedback for the evaluatee nor as a means of improving processes in the project.

The area of evaluation developed because research could not supply satisfactory answers to the kinds of questions that evaluation deals with. The pressure on evaluators and on evaluation to carry out activities using research criteria prevents evaluators from producing satisfactory evaluations.

This article looks at the chief differences between research and evaluation of various types. I have tried to show that not only are the criteria of research unsuited to the demands of evaluation but they also lower its quality. Acknowledging evaluation as an independent discipline, different from research, will enable evaluators to develop their own criteria and appropriate methodology.

NOTES

1. Evaluation examines social, agriculture, community, health, and educational projects; educational syllabi; limited or widespread, simple or complex programs; short- or long-term follow-up. It examines structure, process, relationships, and products at the individual level

or at the level of the system. For convenience, in this article I will refer to all these as a “project.”

2. We are not speaking of “grounded theory” (Glaser & Strauss, 1967), despite an apparent similarity. Although that theoretical approach is anchored in the field, it sees the creation of a theory as the important aim. The difference between this and other research approaches is in direction and systems. In evaluation, the development of a theory is not the main goal.
3. I know the claim that “everything can be measured as an operational variable, even love,” and I agree that the possibility exists. The question is, for what purpose? We can transfer the nature of the neighbourhood in which a project operates into operational variables, but unless we want to compare it to other neighbourhoods, why should we do so? The description of the neighbourhood as a single phenomenon in a project, and its influence, give us all we need in order to understand the project situation.
4. The contract is of great importance in an evaluation. There is a formal contract and an informal contract, which is usually not written (or even formulated orally as a formal statement), between the evaluator and the people involved in the project. We are speaking of the “understandings” that the evaluator and the people being evaluated have about the boundaries of the evaluator’s intrusion and on the ways this will be done. Will the evaluator be permitted to be present at a personal guidance session between the guidance counsellor and a paraprofessional? Are management meetings open to the evaluator or out of bounds? Does the evaluator present his/her findings to the entire staff, or only to management? Are findings presented orally or in writing?

REFERENCES

- Abma, T.A. (1997). Playing with/in plurality: Revitalizing realities and relationships in Rotterdam. *Evaluation*, 3(1), 25–48.
- Ashkenazi, M. (1986). *Research methods in social sciences* (Unit 4: Ethnographic research in anthropology). Tel-Aviv: The Open University. [Hebrew]
- Bechler, Z. (1987). *Philosophy of science*. Tel-Aviv: University on the Air, Israel Ministry of Defense. [Hebrew]

- Brisolara, S. (1998). The history of participatory evaluation and current debates in the field. In E. Whitmore (Ed.), *Understanding and practicing participatory evaluation: New Directions for Evaluation*, 80 (pp. 25–41). Hoboken, NJ: Jossey-Bass.
- Campbell, D.T. (1986). Relabeling internal and external validity for applied social sciences. *New Directions for Program Evaluation*, 31, 67–77.
- Caws, P. (1965). *The philosophy of science*. Princeton: D. Van Nostrand.
- Chen, H. (1990). *Theory-driven evaluation*. Newbury Park, CA: Sage.
- Chen, H., & Rossi, P. (1981). The multi-goal theory-driven approach to evaluation: A model linking basic and applied social sciences. In H. Freeman & M. Soloman (Eds.), *Evaluation studies review annual*, 6 (pp. 38–54). Beverly Hills, CA: Sage.
- Chen, H.T., & Rossi, P.H. (1992). Introduction: Integrating theory into evaluation practice. In H. Chen & P.H. Rossi (Eds.), *Using theory to improve program and policy evaluation* (pp. 1–11). Westport, CT: Greenwood Publishing Group.
- Cook, T., & Campbell, D.T. (1976). The design and conduct of quasi-experiments and true experiments in field settings. In M. Dunnette (Ed.), *Handbook of industrial and organizational psychology*. Chicago: Rand McNally.
- Cook, T., & Campbell, D.T. (1979). *Quasi-experiments: Design and analysis for field settings*. Chicago: Rand McNally.
- Copi, I.M. (1961). *Introduction to logic*. London: Macmillan.
- Copi, I.M., & Burgess-Jackson, K. (1995). *Informal logic*. Upper Saddle River, NJ: Prentice Hall.
- Cordray, D.S. (1986). Quasi-experimental analysis: A mixture of methods and judgment. *New Directions for Program Evaluation*, 31, 9–28.
- Cousins, J.B., & Whitmore, E. (1998). Framing participatory evaluation. In E. Whitmore (Ed.), *Understanding and practicing participatory evaluation: New Directions for Evaluation*, 80, 5–23. Hoboken, NJ: Jossey-Bass.

- Creswell, J.W. (1998). *Qualitative inquiry and research design: Choosing among five traditions*. Thousand Oaks, CA: Sage.
- Cronbach, L.J. (1963). Course improvement through evaluation. *Teachers College Record*, 64, 672–683.
- Cronbach, L.J. (1983). *Designing evaluation of educational and social programs*. San Francisco: Jossey-Bass.
- Drewett, A. (1997). Evaluation and consultation: Learning the lessons of user involvement. *Evaluation: The International Journal of Theory, Research and Practice*, 3(2), 189–204.
- Eyken, W.V., Goulden, D., & Crossley, M. (1995). Evaluating educational reform in a small state: A case study of Belize, Central America. *Evaluation*, 1(1), 33–44.
- Fetterman, D.M. (1996). Empowerment evaluation: An introduction to theory and practice. In D.M. Fetterman, S.J. Kaffarian, & A. Wandersman (Eds.), *Empowerment evaluation knowledge and tools for self-assessment & accountability*. Thousand Oaks, CA: Sage.
- Finne, H., Levin, M., & Nilssen, T. (1995). Trailing research: A model for useful program evaluation. *Evaluation*, 1(1), 11–32.
- Geertz, C. (1973). *The interpretation of cultures*. New York: Basic Books.
- Glaser, B., & Strauss, A.L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Chicago: Aldine.
- Goetz, J.P., & LeCompte, M.D. (1984). *Ethnography and qualitative design in educational research*. London: Academic Press.
- Guba, E.G., & Lincoln, Y.S. (1989). *Fourth generation evaluation*. London: Sage.
- Hempel, C.G. (1965). *Aspects of scientific explanation and other essays in the philosophy of science*. New York: The Free Press.
- House, E.R. (1980). *Evaluating with validity*. Beverly Hills, CA: Sage.
- James, P.D. (1989). *An unsuitable job for a woman*. London: Penguin Books.

- Joint Committee on Standards for Educational Evaluation. (1981). *Standards for evaluations of educational programs, projects and materials*. New York: McGraw-Hill.
- Kazi, M.A.F. (1996). Single-case evaluation in the public sector using a combination of approaches. *Evaluation: The International Journal of Theory, Research and Practice*, 2(1), 85–97.
- Kerlinger, F.N. (1972). *Foundations of behavioral research*. New York: Holt, Rinehart & Winston.
- Kincheloe, J.L., & McLaren, P. (2000). Rethinking critical theory and qualitative research. In N.K. Denzin & Y.S. Lincoln (Eds.), *Handbook of qualitative research*, 2nd ed. (p. 282). Thousand Oaks, CA: Sage.
- Krathwohl, D.R. (1982). *Social and behavioral science research: A framework for conceptualizing, implementing and evaluating research studies*. San Francisco: Jossey-Bass.
- Kuhn, H. (1962). *The structure of scientific revolutions*. Chicago: University of Chicago Press.
- LeCompte, M.D., & Preissle, J. (1993). *Ethnography and qualitative design in educational research*. San Diego: Academic Press.
- Levin-Rozalis, M. (1987). *The process of partially evaluating the change caused by intervention in a community*. Master's thesis in humanities and social sciences. Ben-Gurion University of the Negev. [Hebrew]
- Levin-Rozalis, M. (1998). Are evaluation and research really the same thing? *Megamot*, 39(3), 303–319. [Hebrew]
- Levin-Rozalis, M. (2000). Abduction, a logical criterion for program evaluation. *Evaluation: The International Journal of Theory, Research and Practice*, 6(4), 411–428.
- Levin-Rozalis, M., & Bar-On, N. (1994). *Diagnosis and evaluation in youth village "Eshel-Hanasi"*. Internal report submitted to the youth village board, management, and staff. [Hebrew]
- Lincoln, Y.S., & Guba, E.G. (1986, June). But is it rigorous? Trustworthiness and authenticity in naturalistic evaluation. *New Directions for Program Evaluation*, 30, 73–81.

- Mor, Y. (1992, November). *The use of grey documentation in the comprehensive evaluation of a broadly aimed social program*. Paper presented at the International Expert Meeting on Information and Social Welfare Policy and Planning, Jerusalem.
- Nachemias, D., & Nachemias, H. (1982). *Research methods in social sciences*. Tel-Aviv: Am-Oved. [Hebrew]
- Nevo, D. (1989). *Useful evaluation: Evaluating educational and social projects*. Tel-Aviv: Massada Ltd. [Hebrew]
- Owen, M.J., & Rogers, P.J. (1999). *Program evaluation: Forms and approaches*. London: Sage.
- Patton, M.Q. (1981). *Creative evaluation*. Beverly Hills, CA: Sage.
- Patton, M.Q. (1997). *Utilization focused evaluation*, 3rd ed. Thousand Oaks, CA: Sage.
- Pawson, R. (1996). Three steps to constructivist heaven. *Evaluation*, 2(2), 213–219.
- Pawson, R., & Tilley, N. (1997). *Realistic evaluation*. London: Sage.
- Peirce, C.S. (1960). *Collected papers of Charles Sanders Peirce* (Vols. III, IV) (C. Hartshorne & P. Weiss, Eds.). Cambridge, MA: Belknap Press.
- Peres, Y., & Yatsiv, G. (1994). *Introduction to research methods in behavioral sciences*, 2nd ed. Jerusalem: Academon Press. [Hebrew]
- Popper, K.R. (1959). *The logic of scientific discovery*. London: Hutchinson.
- Renzulli, J.S. (1975). *A guidebook for evaluation programs for the gifted and talented*. Ventura, CA: Office of the Ventura County Superintendent of Schools.
- Rossi, P.H., & Freeman, H.E. (1982). *Evaluation: Systemic approach*. Beverly Hills, CA: Sage.
- Rossi, P.H., & Wright, J.D. (1984). Evaluation research: An assessment. *Annual Review of Sociology*, 10, 331–352.
- Sabar, N. (1990). *Qualitative research in teaching and learning*. Tel Aviv: Massada Press. [Hebrew]

- Schwandt, T.A. (2002, October 12). *Democratization, globalization, and performativity: Implications for learning in and from evaluation*. Keynote address given at the 5th biennial meeting of the European Evaluation Society, Seville, Spain.
- Scriven, M. (1967). The methodology of evaluation. In R.E. Stake (Ed.), *Curriculum evaluation* (AERA Monograph Series on Evaluation, No. 1) (pp. 39–83). Chicago: Rand McNally.
- Scriven, M. (1983). The evaluation taboo. *New Directions for Program Evaluation*, 19, 75–82.
- Scriven, M. (1991). *Evaluation thesaurus*. Newbury Park, CA: Sage.
- Shank, G., & Cunningham, D.J. (1996). *Modelling the six modes of Peircean abduction for educational purposes*. Available online at <<http://www.cs.indiana.edu/event/maics96/Proceedings/shank.html>>.
- Stake, R.E. (1969). Language, rationality and assessment. In H.W. Beaty (Ed.), *Improving educational assessment and an inventory of measures of affective behavior* (pp. 14–40). Washington, DC: Association of Supervision and Curriculum Development.
- Stufflebeam, D.L. (1971) The CIPP model for program evaluation. In G.E. Madaus, M. Scriven, & D.L. Stufflebeam (Eds.), *Evaluation models: Viewpoints on educational and human services evaluation*. Boston: Kluwer-Nijhoff.
- Stufflebeam, D.L., Foley, W.J., Gephard, W.J., Guba, E.G., Hammond, R.L., Merriman, H.O., & Provus, M.M. (1974). *Educational evaluation and decision making*, 4th ed. Itaska, IL: F.E. Peacock.
- Tucker, A.S., & Dempsey, J.V. (1991). A semiotic model for program evaluation. *The American Journal of Semiotics*, 8(4), 73–103.
- Turner, J.H. (1986). *The structure of sociological theory*, 4th ed. New York: Wadsworth.
- Tyler, R.W. (1942). General statements on evaluation. *Journal of Educational Research*, 35(7), 492–501.
- Tyler, R.W. (1950). *Basic principles of curriculum and instruction*. Chicago: University of Chicago Press.

Wallace, L.W. (1971). *Sociological theory*. London: Heinmann Educational Books.

Weiss, C.H. (1996). Theory based evaluation: Past, present and future. Paper presented at the annual meeting of the American Evaluation Association, Atlanta, GA.

Miri Levin-Rozalis, Ph.D., sociologist and psychologist, is a faculty member of the Department of Education at the Ben Gurion University in Beer-Sheva, Israel. She has been an evaluator of projects programs and services in the social and educational fields and an organizational consultant within the university and as a freelance for more than 20 years. Until June 2003 she was the president of the Israeli Association for Program Evaluation (IAPE) and is still a member of the executive management.

