THE LIMITS AND THE SPECIFIC INSTRUMENTS OF POLICY EVALUATION[1]

Cosmin BRICIU[2]

Abstract: *This paper offers arguments that policy evaluation should be considered and researched as a distinctive field with its own challenges and limits. The paper puts forthsome of the approaches that might be useful in policy evaluation endeavors. In order to achieve this task, the definition and some of the most important concepts of evaluation are reviewed first, using a variety of sources. The concluding discussion emphasizes some of the directions for research in the area of policy evaluation, singling out strategies as a useful policy tool for the use of policy evaluation attempts.*

Keywords: *policy evaluation, literature review, theory of intervention, policy cycle.*

Policy evaluation– definitions and concepts

This paperdeals with policy evaluation, regarded as a distinct and legitimate area of evaluation. The current literature on the subject, especially the academic one, concerned with evaluation theories, methods and techniques, seldom, if ever, makes this point. Usually, policy evaluation is either regarded as the less respectable relative of program evaluation, as there are too many variables in the process to allow control over them, or it is considered equivalent, explicitly or implicitly, with program evaluationfrom the point of view of the evaluators' tasks.

Policy evaluation is placed at the crossroad of several thematic areas: evaluation (most of the theoretical contributions in this area elaborate on program evaluation methods and techniques), policy/social policy analyses and social development. While scientific contributions authored by evaluators and evaluation theorists on the specific subject of policy evaluation are scarce, policy analyses texts are more generous in this respect,

[2] Researcher, Research Institute for Quality of Life, Romanian Academy, Bucharest, ROMANIA. E-mail: cosminbriciu@gmail. com

marking evaluation as an intrinsic phase of the policy cycle. However, the latter also do not indicate,oftentimes,the specifics of policy analyses and are confined to re-stating the principles, the approaches and the methods of program evaluations; these sources recall the basic distinctions in the evaluation field such as formative versus summative, input, output and outcome, experimental, quasi-experimental and non-experimental and so on. Policy analyses texts tend to be more concerned with the environment and the user end of the evaluation process: they stress that evaluation is a policy process itself and display more poignantlythe difficulties in performing rigorous analyses in the framework of the ever changing and complex policy-making arena. The social development textbooks and guides are closer to the approach taken here, of regarding policies and strategies as interventions that can be and should be measured and evaluated and offer practical advice but they miss some of the valuable points made by program evaluation literature, such as the emphasis placed by the latter on identifying the theory of the intervention prior to proceeding to its evaluation.

Some definitions are needed, not strictly for taxonomy purposes, but in an attempt to restrict the concepts used here. Throughout this paper, the meaning of policy is that of a large intervention, comprising legislative provisions, strategic documents, several programs, institutional set-ups and many other contextual arrangements. It is an intervention wider in scope than programs and projects, as the latter are usually defined (*e. g.* Mățăuan, 1999: 38).

Not all generally accepted definitions of policy point towards this meaning. For instance, Jenkins givesthe following definition: "a set of interrelated decisions taken by a political actor or group of actors concerning the selection of goals and the means of achieving them within a specified situation where these decisions should, in principle, be within the power of those actors to achieve"(1978: 15). Other popular definitions might be: "a policy is a set of ideas or a plan of what to do in particular situations that has been agreed officially by a group of people, a business organization, a government or a political party" (Cambridge dictionaries online). In the widestsense of the concept, policy can mean anything from a mere change in the legislative provisions, for instance the level of taxes, without any other additional arrangements or actions, to the strategy of a certain governmental cycle.

Most of the times, the current definitions of evaluation refer to program evaluation. For instance, perhaps the most widely read and quoted evaluation textbook available today states that «evaluation research is the systematic application of social research procedures for assessing the conceptualization, design, implementation, and utility of social intervention programs (Rossi, Freeman &Lipsey, 1993: 5). Another definition states "program evaluation is the systematic collection of information about the activities, characteristics, and outcomes of programs for use by specific people to reduce uncertainties, improve effectiveness, and make decisions with regard to what those programs are doing and affecting". (Patton, 1986: 14)

The perspective on "evaluation as a phase of the policy cycle" is closer to the approach here. However, even in the framework of this perspective, the process is oftentimes comprised of a monitoring component which pertains to the policy as a whole, and evaluations which pertain only to particular programs; the information generated by the

analysis is construed as a phase of the policy cycle. Wollman (2007: 393) offers a more comprehensive definition: "first, evaluation research, as an analytical tool, involves investigating a policy/program to obtain all information pertinent to the assessment of its performances, both process and result; second, evaluation as a phase of the policy cycle more generally refers to the reporting of such information back to the policy cycle" (Wollman 2007: 393). There are other definitions as well which make this distinction: "evaluation is the systematic assessment of the operation and /or outcome of a program or policy, compared to a set of explicit or implicit standards, as a means of contributing to the improvementof the program or policy" (Weiss, 1998: 4). The same author shows that "evaluation is a type of policy research, designed to help people make wise choices about future programing; evaluation does not aim to replace decision makers'experience and judgment, but rather offers systematic evidence that informs experience and judgment" (Weiss, 1986: 83).

However, even when the definitions cover policy as well, in the practical sections wherethe evaluation operations are described, it becomes evidentthat the definitions, in fact, refer to programs, or that no specific difference is considered. The theoretic and methodological problems are expected to be more or less the same when dealing with policies, programs or projects as long as there is no specific guidance on how policies should be tackled by the evaluator.

For the purpose of this paper, Scriven's vision on evaluation is more helpful. For Scriven, evaluation is used in the judiciary process (as a central element of trials), in engineering, in medicine, in logic or in product evaluation, to name only a few instances where the evaluative dimension is a central component. As such, Scriven states that evaluation is a *transdiscipline* that "supplies essential tools for other disciplines, while retaining an autonomous structure and research effort on their own" (Donaldson, Scriven, 2009: 19). This view contrasts with the already mentioned definition of Rossi, Freeman and Lipsey's volume, claiming that evaluation is the application of social science methods to the solution of social problems. It's not just that "evaluation is an elastic word that stretches to cover judgments of many kinds" (Weiss, 1998: 3), it is the fact that these judgments are carried on with a fairly high level of similarityacross the fields, at least in respect with the intention and scope of the endeavor.

The communalities are even stronger, as one might expect, between policies and programs. However, there are also specific features and they are the subject of interestof this paper. While authors brought up in the tradition of random test evaluation techniques are cautious to take on directly the problem of policy evaluations, in the social analyses field the problem is tackled: "the concept of policy evaluation thus refers broadly to the stage of the policy process at which is determined how a public policy has actually fared in action" (Howlett, Ramesh, 2003: 207); "policy evaluation assesses the effectiveness of a public policy in terms of its perceived intentions and results" (Gerston, 1997: 20).

Having summarily inventoried some of the definitions of evaluation and policy evaluation, is there something to say for the case of policy evaluation as a substantive domain of evaluation? We can observemacro-level analyses that are actually carried on and that involve a strong evaluative dimension, even though they are not labeled as

such. These analyses, not commonly referred to as evaluations, have some of the main characteristics as the latter: they are normative and ascertain that policies were right or wrong in their intent and application and they do indeed work with hypothetical theoretic models of causality (which they do not entirely demonstrate, however). Scientists as well as high-level officialsassess performances of the social policies in their integrality on regular basis.

The paper builds on the idea that it is important to better ground, at the theoretical and methodological level, such macro-level analyses and others of the same kind,as we might be interested in taking decisions not just at the level of individual programs but strategic decisions concerning the future of our societies and their policies. In fact, it is a rare case when we can isolate a program from its context, since at least "big problems tend to have a lot of 'solutions' thrown at them, making it difficult to asses which, if any, of them are producing an effect. On the other hand, some programs may work only in conjunction with others, so a research strategy to try to determine the effects of such programs in isolation would be counter-productive." (Hogwood, Gunn, 1984: 227)

In the context of these arguments, we will proceed with the analysis based on the definition that evaluation is "the periodic analysis of the relevance of a strategy, of a plan, of a program or project, of their efficiency and impact, of the intended and unintended effects, in relation with the established objectives. The evaluation can lead to adjustments precisely because it is focused on the analyses of the global efficiency" (Zamfir, Stoica, Stănculescu, 2007: 80)

Challenges and specific endeavors for policy evaluation

As already stated, policy and program evaluation are quite similar in some respects. First and foremost, the intent and scope of the evaluator work is the same, namely to distinguish good interventions from bad ones. This is achieved by comparing the actual state of the policy/program with a desirable state, which is what evaluations in general do: "Evaluation tends to compare *what is*with *what should be*" (Weiss, 1998: 15). However, at this point a series of difficulties emerge:

- While at the level of programs, the "should be" part can be quite often extracted from the program goals and objectives or can be reconstructed through alternative strategies designed to isolate the so-called "theory of the program", the "should be" component of the policies is less obvious;

- Overall, it is quite uncommon that a nuanced analyses of policies would conclude that they are good or bad overall, not totally unlike social programs but even more so; the evaluation of policies differentiatesmore among policy components, such as legislations, programs activities and so on.

Understanding the theory of thepolicies and programs

Having said that, a quick review of the discussions of the evaluator theorists on the "program theory" reveals that there are considerable difficulties in identifying this component at the level of programs as well.

Program theory comprises the "*what should be*" component at the level a particular program (and policy) and also shows "*how is this desirable goal supposed to be achieved*". "Although the term *program theory* often seems to conjure up images of broad social science theories about social problems, theory-driven program evaluators use the term to refer to rather small and specific theories of social programs, treatments, or interventions" (Donaldson, 2009:130). The theory-driven evaluation is "a plausible and sensible model of how the program will work under certain conditions to solve identified problems" (Bickman, 1989: 5).

The model is not confined to presumed cause-effect relation but it also details all the characteristic of the particular setting where a program is implemented: "Theory-driven evaluation provides a feasible alternative to the traditional method-driven approaches that have come up short in (…) (reporting – author's note) on which program components are the most effective, the mediating causal processes through which they work, and the characteristics of the participants, service providers, setting and the like that moderate relationships between a program and its outcomes" (Donaldson, 2009: 114).

The theory-driven evaluation is considered by many evaluation theorists to be superior to method-driven approaches for the very reason that it can determine the contextual use of the methods for particular programs in particular settings: "Evaluation-theory tells us when, where, and why some methods should be applied and others not, suggesting sequences in which methods could be applied and others not, ways different methods can be combined, types of questions answered better of less well by a particular method, and benefits to be expected from some methods as opposed to others" (Shadish, 1991:34).

Scriven (2009: 22-23) makes the distinction among levels of the program theory:

- the "*allegedprogram theory*" or the official version how a program operates or is believed to operate by the decision-makers, funders and all the stakeholders who support the program; Scriven emphasizes that it is not identical with the goals of the program (but the goals are an important component, one might add)

- the "*real logic of the program*", the actual mechanism that allows the implementation of the program; it is familiar among the practitioners or might be reconstructed from their partial knowledge

- the"*optimal program theo*ry" or the account of how the program *should* operate, in order to achieve optimal effectiveness from available resources

Weiss makes a distinction between the *program theory*and *implementation theory*. The former, he submits, represents "the set of beliefs that underlie action" (Weiss, 1998: 55) (which are not supposed to be uniformly accepted or right; they are set a hypotheses

upon which people build their program plans);, while the latter is"a theory about what is required to translate objectives into ongoing service delivery and program operation; (…) the assumption is that if the activities are conducted as planned, with sufficient quality, intensity, and fidelity to plan, the intended results will be forthcoming" (Weiss, 1998: 58).

In order to understand the theory of the programs /policies, the first step is to identify the objectives. However, the difficulties in identifying the objectives are considerable: "any emphasis on examining the extent to which policy objectives are accomplished by a program must contend with the reality that policies often do not state their objectives precisely enough to permit rigorous analyses of whether they are being achieved" (Howlett, Ramesh, 2003: 213).

Considering the widely acknowledged difficulties in extracting the program theory, various researchers have elaborated a series of guidance notes on how to conduct research. These guidelines are even more useful when implemented in order to understand the theory of policies, which are sometimes more elusive than the programs. In respect with the identification of goals, the researcher might:

- Draw up the statement of goals himself, after reading the program documentation, talking to practitioners and observing the program in action

- Set up a collaborative effort in formulating goals, sit down with practitioners and discuss and refine alternatives

- Conduct an exploratory open ended study instead of a formal evaluation with the question of goals left aside

- The researcher might ask members of the target group (Hogwood, Gunn, 1984: 223)

In order to understand the theory of the program beyond its alleged or real objectives, and grasp a meaning of the logic ofintervention behind the actions, similar and other complementary strategies should be put to work (Leeuw, F. L., 1991: 22-23):

- Search in policy documents for statements that deal with the problem or the problems in review; apply methods of content analyses

- Specify stakeholders who can be defined as policy makers; describe their role

- Search for reasons why it is believed necessary to solve a certain problem and why it is believed to solve it by way of policy

- Make an overview of the statements of stakeholders concerning the goals of the policy; reformulate these statements in terms of "if-then propositions"

- Bring these propositions together in a system of propositions and search for missing links and inconsistencies

- Go into detail with respect to missing links, apply "argumentation analyses" and interview stakeholders again in respect with these missing links: argumentation analyses should focus on warrants; a warrant is the "because" part of an argument

- Reformulate these warrants in terms of "if then" and add them to the system, to be characterized as a policy theory

- Apply triple-theory assessment: the policy theory is evaluated from three perspectives: first, the epistemological one (one wants to find out to what extent the policy theory is in line with results from scientific research); second, the theory is assessed from the perspective of policy feasibility; and third, the theory is assessed from the perspective of the manipulability of the main variables included

Despite all these operations, it might prove difficult to pin down the theory behind certain policies, especially after their implementation and especially the one that Scriven called alleged theory. It is difficult enough to establish basic objectives for future strategies, without the haziness added by the passage of time. In order to be implemented, policies require the consent of several high-level officials (the decision-makers within the sector in question, the finance decision-makers and so on). They might be the result of a favorable convergence of efforts from various directions and with slightly different purposes and "theories".

However, when the strategies /policies are not just an inventory of desirable actions and they are really conceived envisaging a different future for the societies, they resemble programs in respect with the fact that they propose a change of some "strategic variables" (Zamfir, 2007: 37) that would trigger social development. Although the mechanic of the hypothetic causality that they project on the foreseeable future is more complex to be described than in the case of programs, consistent policies and programs should be formulated on the basis of a theory of change. However, this theory of change is admittedly more fragmented than in the case of programs, as it might envision various types of new developments, not always closely interrelated with each other.

Another challenge might be the fact that the interested parties are rarely interested in inquiring about the value of large interventions during long periods of time: "the decision-makers are under the pressure of the budgetary cycles and they want fast answers to the questions which, most of the times, do not adress long periods of time" (Cace, 2002: 25)

The problem of causality

As already mentioned, the problem of causality is central in evaluation, and it determines the preference of many evaluation practitioners for random selection design programs. "The intervention is only one of many influences on the target problem, because unforeseen changes may arise, because the program may interact with a number of other programs, and because it may be difficult to separate out the effects of a "new" program from the long-term effects of programs which it replaces." (Hogwood, Gunn, 1984: 20)

The highest technical standard in the evaluation field, which is to have randomized trials to test the impact of programs, becomes useless in the case of policy evaluation, since policies makerscannot afford to implement policies addressing only partsof the

eligible population. However, as already mentioned, causal relationships are envisioned in policies as well, although frequently in a looser and more difficult to assessmanner.

The roles of key stakeholder consultation and participation of the beneficiaries are critical

The increasing sense of importance of the involvement of stakeholders and beneficiaries in the evaluation process has determined the design of approaches centered on this distinctive methodological feature, such as "inclusive evaluation" (Mertens, 1999) or "empowerment evaluation" (Fetterman, 2000). The premise of these perspectives is that knowledge itself is not neutral but reflects the existing social inequalities and power imbalances. The judgmental character of evaluation makes it prone to reinforce the existing establishment unless it is carefully designed to prevent this risk. These approaches use methods such as self-evaluation, self-reflection of the beneficiaries and encourage program participants to take ownership of the evaluation process and findings. They can employ a variety of methods and techniques although "they can be missinterpreted as being qualitative methods" (Bleahu, 2004: 17).

The supporters of these methods stress the importance of reflectionon the values that are promoted: "evaluators must negotiate whose questions will be addressed and whose interests will be served by their work (Greene, 1994: 531).

At the level of policies, the range of stakeholders and types of beneficiaries is even higher than in the case of program evaluation, which makes the process difficult but not less needed.

Policy evaluations are subject to contestation and the conflict of different views and perspectives is sharp

Despite efforts to secure the participation of all stakeholders and participants to the process, policy evaluationis considered by some theorists to be a political process itself; hence, opposing views on such diverse issues as the nature of the social problem, the required intervention and the expectations towards the evaluation endeavor are foreseeable.

Another reason for debate is that evaluations ascertain cause-effect relations to policies, as already mentioned: they also discern intentionality behind policies and discuss and theories behind actions. However, for all the above reasons, they are more liable to be contested and they are contested on regular basis by other researchers.

It is recommendable that some form of evaluation is performedbeforethe policy design and implementation phases

There are two potential uses of evaluation in the inception phase of policies and programs. First, a monitoring and evaluation system should be built in policies and programs in order to support tracking the progress and ensure accountability. This has become common practice across many governments and non-governmental organizations throughout the world. Generally, the standard to elaborate monitoring

systems goes undisputed at the program and at the policy level. However, pre-plannedevaluations are controversial. At the level of programs, there are pros and cons regarding the establishment of pre-planned evaluation systems. For instance, they might influence the behavior of the participants and "the pre-plannedmethod may frequently take on some of the characteristicsof any longitudinal study and on balance would incur larger costs than a one shot post-planned evaluation" (Strasser, Deniston, 1978: 195). Second, a pre-evaluation of the expected effects should be attempted: "front-end analyses refer to evaluations undertaken before an intervention is introduced or a program adopted" (Greene, 1994: 13). The high level of resources involved in the implementation of a policy should determine the generalization of the practice of pre-evaluating what are the expected results. Usually, at the government level, some sort of pre-evaluation is carried on, at least a review of the exiting evidence on relevant programs and services or some sort of statistical modelingof the expected effects. When the monitoring systems are under-developed, an initial evaluability research can be conducted (Annex Table 1) to find if regular collection, reporting and analyses on performance-based management and evaluation (M&E) systems can be insured. Bochel and Duncan (Annex Table 1) propose a quite generous list of evaluation options available for future interventions: basic research, policy analyses, prototypes, micro-simulation, program evaluation (impact or summative evaluation), random assignment, matched designs, cohort designs, statistical controls, prospective evaluation, laboratory experimentation and gaming (depending on the timing of the future – close future, future or expansive future).

There is another way to consider the logical and chronological relation between interventions and evaluation. According tosome theories, social problems are acknowledged and tackled only when the stakeholders or society have already identifiedavailable and feasible solutions. Up to that point, problems are "latent" (Cătălin Zamfir, 1977). This theory suggests that a rough pre-evaluation of efficiency based on loose concepts and arguments is performed from the very moment when a certain condition is defined as a social problem.

There is one final argument to be made about the necessity of an initial evaluation and it concerns the utility of research: "the failure of evaluation research to feed significantly and successfully into the policy process may be explained via a stunningly obvious point about the timing of research vis-à-vis policy – namely, that in order to inform policy, the research must come before the policy" (Pawson, 2009: 8).

Policy learning

The process of policy evaluations is a good illustration ofan idea that hasgainedgradual acceptance in the area of program evaluation area as well, with the early rationalist models, oriented towardhard facts finding,leaving room for more recent models emphasizing that policy evaluation is continuous and less than perfect policy learning. Early approaches took an optimist hue, and policy evaluation was believed to represent "the objective systematic, empirical examination of the effects ongoing policies and public programs have on their targets in terms of the goals they are ment to achieve" (Howlett, Ramesh, 2003: 219). However, with the maturation of the field, and with

growing numbers of evaluation finding nil or modest effects for the program analyzed, this initial ambition was questioned. Some theorists now claim that "perhaps the greatest benefit of policy evaluation is not the direct results it generates, but the educational process it can engender" (Pressman and Wildavsky, 1984).

The shift from the positivist scientific paradigm to more pragmatic evaluation postures is usually illustrated in the evaluation literature by the positions expressed by two major theorists, Donald Campbell and Lee J. Cronbach. Campbell (1969) was envisioningan experimental society, with policy decisions emerging from continual social experimentations that test solutionsfor improvement of the social conditions. This is a quite similar vision to that of scientific knowledge in general: in this case, the programs are hypothesis and the evaluations provide the test. In Cronbach's vision (1982), evaluation is more art than science and every evaluation should be tailored to meet the needs of decision-makers and other stakeholders. Thus, where scientific studies strive to meet research standards, evaluations should be dedicated to providing maximally useful information for decision makers given the political circumstances, program constraints, and available resources.

However, the conceptthat programs and policies provide input for the adjustment of interventions has also been contested and considered a simplified model of the actual process. Rist suggests that evaluation findings serve more asconceptual inputs rather than instrumental ones. In the first place, it is rarely the case that an evaluation explicitly indicates what solutions should be adopted: "Evaluation rarely points in an unequivocal policy direction" (Bochel, Duncan, 2007: 177). In the second place, it is even less common that the decision-makers adopt the recommended solutions. Both arguments are stronger in the case of policies, since even when some options are foundundesirable by a pre-evaluation, the remaining policy options are too numerous for the experts to point a single recommended course of action. "The instrumental use suggests direct linkage and application of the evaluation findings and recommendations to decision making. Conceptual use suggests an indirect application whereby the policy maker begins to think differently about the problem or condition, frames the possible policy approaches in a new manner, and perhaps anticipates a different set of outcomes." (Rist, 1995: xiiv). The findings of evaluation feed in the policy process as bits and pieces of knowledge, filtered by factors such as limited receptivity of the decision-makers, limited flexibility of the institutional set-up, the different nature of the interests of the research and decision-making side, and the limited usefulness of the evaluation findings for decision making processes. "Studies indicate that evaluation research is used conceptually much more often than it is instrumentally" (Albaek, 1990: 9) (…) in unsystematic and diffuse ways, social science findings, including data, concepts and theories, reach decision makers" (Albaek, 1990: 8).

Given this filtered relation between evaluation and research, the ambitions of the evaluators to influence the policy-making process should remain modest: "decision making in the policy arena can be broadly characterized as occurring at two levels. The first level involves the establishing of the broad parameter of governmental action, e. g. providing national health insurance, restructuring national immigration laws etc. At this level and in these instances, policy research input is likely to be quite small, if not nil (…) Once the issue is on the agenda of key actors or organizations within the policy

establishment, there are possibilities for the introduction and utilization of policy research and utilization" (Rist, 1995: xviii).

Nevertheless, instrumental criteria should be maintained as a standard for evaluations, although the actual use of evaluation is often merelyconceptual: "the emphasis in assessing the strengths and weaknesses of policy evaluation is not and should not be on methods. Methodological techniques, regardless of how elegant or comprehensive, are not the yardstick by which the field is to be judged. Rather, the judgment should be made on the degree to which evaluation has demonstrated the consequences of present and past policy initiatives, clarified present policy choices, informed decision makers as to the costs and benefits of different options, and re-framed the debates on pressing national problems" (Rist, 1995: xiiv)

There is also a considerable risk in placing too strong an emphasis on the inherent political character of evaluation, whether the acknowledgment of the strive of evaluations to serve as instruments for decision-making deters evaluators from seeking valid conclusions based on program performances. In such cases, oftentimes, "the policy-research relationship is financially circular, with one arm of government providing the funds for another to supply the evidence base" (Pawson, 2009: 3). These type of "politically oriented studies" are labeled as "pseudo-evaluations" by theorists (Stufflebeam, 2001: 91)

Forss, Rebien and Carlsson (2002) have identified five types of utility of the evaluations: as a learning process; as a means to develop professional networks; for strengthening mutual communication and understanding; as a means to consolidate the project; as a means for moral boosting (Forss, Rebien andCarlsson, apudNeguţ, Nicolăescu, Preoteasa, Cace, 2011: 61)

Policy evaluation is continuous: the culture of evaluation

Using the classical distinction between formative and summative programs, policy evaluation is formative by its nature, aiming to inform the ongoing policy process. The multi-layered, continuous and changing nature of the policy implementation most of the times calls for process-oriented evaluations instead of results-based assessments. This is yet another methodological consequence of the non--experimental feature of the policy implementation process than the already noted difficulty in controlling the outcomes via randomly assigned control groups. Performing a process-oriented evaluation means assessing the policy "with the premise that it is important to learn how a program actually works, and what are it weak and strong points. This type of evaluation is useful especially in the case of the programs carried on for lengthy periods of time, which have changed during their implementation" (Preoteasa, 2004: 53).

The formative character of policy evaluation matches the ongoing nature of policy implementation and perpetual re-emergence in the society of most social problems. Another term used to describe this approach is "constructive evaluation" (e.g.: Palumbo, Hallet, 1993). "There are two very distinct approaches to thinking about policy practice. One is teleological, outcome-focused: the activity is about "making policy", and the focus of attention is on the problem being addressed and how the

measures proposed would contribute to its solution. The alternative approach might be termed "relational" or process focused: policy activity is a continuing but variable flow of attention among a large and diverse ray of participants, who have overlapping agendas, different interpretations of the problem, and varying levels of concern about its resolution." (Hoppe, Noordegraaf, 2010: 228)

A central role for evaluation is to create a culture of evaluation across policy fields: "while evaluation aims to establish whether a policy works, it has a broader, more sustained purpose that can be beneficial even if the results of the evaluation itself are inconclusive. It requires policy makers to be specific about the objectives of policy, to express them in ways amenable to measurement and, ideally, to assign objectives different priorities" (Walker, Duncan, 2007: 172).

Specific tasks for the policy evaluation endeavor

Why should one be concerned with the larger picture, *i.e.* policies and their monitoring and evaluation? It may be useful to recall two fundamentalarguments:

- The most important social problems are usually tackled by policies, such as strategies, with their corresponding implementation plans, social services, programs and all the encompassing legislative and institutional set-up; for instance, it is not evident that when assemblingtwo programs with certified efficiency in a single policy action, their positive effects will be preserved. It is arguable that such concerns should become more prominent, as "the evaluation literature has become too obsessed with the finer details of methodological improvement at the expense of recalling the purposes of evaluation, and the political context of evaluation" (Hogwood, Gunn, 1984: 228).

- There is also a practical argument to be made, *i. e.* programs are tangled in such a complex way in the web of policies that evaluations can rarely isolate the activities that comprise a program, in order to see it as a black-box with inputs on one side and outputs at the other: "a particular program may be only one of the many tasks to be carried out within a department and may not be allocated a separate section within it. This may be true even at the level of the individual, say, social worker, who will typically be expected to implement a wide range of specific policies. It may be often be difficult to draw boundaries round a set of activities and outputs that constitute of problem". In some cases, it might make more sense to evaluate a policy overall (Hogwood, Gunn, 1984: 221)

The view here is that "interventions are complex systems thrust amidst complex systems" (Pawson, 2009: 168).

Build and capitalize on the cumulative character of evaluation

In the area of program evaluation, there is an ongoing debate about the need to strengthen the cumulative character of findings of various evaluations. However, these discussions revolve around the methods to build through meta-analysis techniques some kind of generalizable learning from evaluations concluding on positive returns of

several similar programs in different settings. "Meta-analysis is the systematic summary of the results from a number of evaluations of the same kind of program." (Weiss, 1998: 48)

A so-called protocol, consisting of a set of procedures, is applied to several evaluations, in order to ensure procedural uniformity. The procedures consist of formulating the review question, identifying and collecting the evidence, appraising the quality of the evidence, extracting and processing the data (presenting the raw evidence on a grid gathering the same points of information from all primary inquiries), synthetizing the data and disseminating the findings (Pawson, 2009: p. 41).

"Once a number of quantitative studies have been done, evaluating the same kind of program, it is possible to combine their results to get better and more generalizable estimate of program effects. The single program (and program evaluation) is the prisoner of its setting. Evaluation results hold for the special place, time, staff, participants, and external conditions that were on the scene when the study was done. Pooling results of many evaluations shows the nature of outcomes of many similar programs across a range of settings" (Weiss, 1998: 236).

However, some theorists consider that this kind of inductive approach does not work in fact: "The bad newsis that the recipe does not really work (…) At every stage of the meta-analyses review simplifications are made. Hypotheses are abridged, studies are dropped, program details are filtered out, contextual information is eliminated, selected findings are utilized, averages are taken (Pawson, 2009: 43).

Even if the only legitimate conclusion reached by a meta-evaluation – less concerned with uniformity – is that certain programs work in certain settings, this is still valuable insight for policy makers. A similar cumulative knowledge might be built around various component themes of a larger evaluation issue, such as poverty and tackling programs, *i. e.* cash transfers, social services, labour market activation and so on. Some of the problems might be common across fields, as would be the case of stigma, for example, while others might be specific for the sector and type of program in question.

The evaluation process is not limited to drawing lessons to be learned but has to verify their utility in the given context. "The process of drawing a lesson involves four analytically distinct stages. The first is searching experience for programs that, in another place or time, appear to have brought satisfaction. Second, it is necessary to abstract a cause-and-effect model from what is observed. The third stage is to create a lesson, that is, a new program for action based on what has been learned elsewhere. Finally, a prospective evaluation is needed to estimate the consequences of adopting a lesson." (Rose, 1993: 27)

Evaluation as part of policy planning

There are obvious links between the desire to ensure that policy is "evidence based" and the need to see what it has achieved, with a natural feedback cycle to subsequent policy improvement (Hill, 2009: 279)

The policy analyses and planning process involves several classic stages that are in their essence evaluative processes, comprising at least the following (they are not necessarily subsequent):

- A needs assessment as a result of the diagnosis of the social situation and social problems

- An evaluation of the intervention needs: some intervention needs can be directly derived from the situation analysis: if there are large numbers of homeless people, an outreach emergency system has to be put in place and some kind of development program has to be figured out, such as the Housing First programs; if there are large numbers of unemployed, activation services are needed; this is the phase when a meta-analyses of the existing evaluation of the exiting services is most useful

- An analyses of the available response at the policy level

- The evaluation of the discrepancy between intervention needs – available response at the policy level

- An analyses of the available resources

- An initial mapping of the alternative solutions and the pre-evaluation of the effects

The difficulty of drafting solutions in the context of scarce resources available in the initial stages is visible in the fact that strategies almost invariably demand more resources than available at the moment. A possible solution for the scarcity of the resources and multitude of efforts required would be the use of "strategic variables", *i.e.* components of the system with two characteristics: (a) a change in their level produces desirable changes within the overall system; (b) they can be changed through direct action (Zamfir, 2007: 37)

The line traditionally drawn between external and internal evaluations is limited from the perspective of the monitoring and evaluation process as a permanent component of the policy cycle. Rist (1995: xix) elaborates on questions that the evaluation has to answer in the stages of the policy formulation, policy implementation and policy accountability.

When the policy is drafted, the diagnosis is the first activity and the first set of information needs regards the understanding of the policy issue. The second set of information needs concerns the previous policies implemented and the knowledge on their effects. Rist emphasizes there is little systematic policy evaluation regarding the intended and unintended consequences of the various policy instruments, which "leaves the policy makers essentially to guess as to the trade-offs between the choice of one tool and another".

At the level of policy implementation, the social problem needs re-examinationdue to its dynamism. There is a need to understand how the policy is internalized at the different layers of the government system (regional, local) and the institutional capacity

to undertake the required effort. A final cluster of information at this level refers to the expertise and qualification of those responsible, the interest of management in careful implementation, the controls in place regarding the allocation of resources, the organizational structure, decision paths to respond to competing demands, disambiguation strategies on the role of the institutions and interactive or feed-back loops.

The evaluator takes on the accountability issue only when the program has reached the maturity so that the effects might be visible (common understating of the evaluation is limited to the analyses undertaken in this stage); the issues of efficiency and impact become central. Accountability, according with the same author, also refers to the quality of the management supervision, leadership of the organization with a clearly articulated vision and goals understood by staff, the attention to processes and procedures that would strengthen the capacity of the organization to implement effectively the policy objective, the use of data-based decision making and the alignment between the leadership and the staff.

Comparing similar contexts /societies

While comparison between a treatment and a control group are almost invariably impossible at the level of policies, this does not mean that the comparative approach is left outside the scope of policy evaluation.

Comparison is implicit in the process of findings lessons for the learning, for instance. Although the comparison cannot benefit from a counterfactual analysis, sometimes the similarities among communities or societies entail analytic endeavors, which are relevant up to a point. For instance, one might compare the profile of social policies within a group of countries belonging to a common geo-political area, such as the former communist central-eastern countries. Another example: we see frequent comparative analyses of the level of public funding of public services in different countries. While the differences among the compared entities are in general to large to account for (and they cannot be usually controlled with statistical procedures, unless an international survey is used), it is nonetheless fruitful to make such analyses.

As far as the outcomes of the policy are concerned, there are three types of evaluations typically carried on in order to assess effectiveness, efficiency and impact. In addition, an evaluation may also analyze the relevance of the policy. "A policy is described as relevant if the objectives are adapted to the nature and temporal and socio-spatial distribution of the problem that the policy is intended to resolve" (Knoepfel, Larrue, Varone, Hill, 2007: 234).

Before and after studies and reflexive controls

In program evaluation, the analyses that use the same group of beneficiaries as a pre-treatment control group and post-treatment experimental group are called reflexive controls. "Reflexive controls is the evaluation strategy assessing changes on outcome measures that occur between the time before targets participate in a program and some point afterward" (Rossi, Freeman, Lipsey, 1999: 343).

By their nature, policies cover the whole target population of a certain geographical area, at least with some of the services and programs that they deliver. The surveys carried on regularly (usually the surveys of the national statistical institutes) might entail reflexive controls analyses of the new policies, usually with some add-ons such as supplementary questions and adjustments of the sample in order to include a representative number of individuals from the target group.

The main methodological limitation is that changes in the outcome variables might have been provoked by other external factors. The eventuality that other variables affected the target group is called the history threat.

Building M&E systems

Establishing M&E systems is the practical way to ensure that some kind of cumulative knowledgeis built at the level of sector policies. The development literature is less concerned with the conceptual difficulties of measuring the progress and assessing the results of policies and more inclined to offer practical advices and tools in order to establish such M&E systems. Ideally, policies should have a permanent M&E system set up to measure their performances, with evaluation being a distinctive component.

"A system is defined as a group of interacting, interrelated, or interdependent elements forming a complex whole and "systems thinking" is about gaining insight into the whole by understanding the linkages and interactions between the elements that comprise the whole system. Applying such a systems approach to M&E systems building, it requires: (i) Identifying the components of the system (understanding that they are interrelated) as a means to describe the system; and (ii) Ensuring that each component is functional to ensure that the system is functional." (Görgens, Kusek, 2009: 7)

Several authors emphasize the strategic and instrumental role of the monitoring systems for policy planning, beyond their function asoutcome measurement instruments:

- Monitoring has the important role of validating (confirming, recognizing outcomes) objectives and resolving differences and ambiguities, especially in face of opposition to change and reform" (Mosse, 2002: 43)

- The monitoring system may be less a means manage the ultimate end of impacts or transactions– either "quantitative" or "qualitative". Rather, the monitoring system must play a major part in establishing the framework of discussions and negotiate the common meanings and resolutions which in turn allow the reinterpretation of positions and the derivation of reassurances about the negotiation and processing of benefit and loss (Rew and Brustinow, 2002: 187).

The evaluation of strategies

Anyone who has ever participated in elaborating strategies, as the author of this paper has had the privilege to do on several occasions, might have remarked that it is not a habit among policy-makers to examine the results of the strategies. Strategies are usually

equipped with monitoring systems comprising indicators of progress and, in the fortunate cases, monitoring reports are issued on a regular basis. At the end of a cycle, a strategy is dropped and another one is formulated.

Another widespread strategy is to choose a set of outcome indicators, such as "the poverty rate" or the "employment rate" as the main monitoring tool. However, as acknowledged here, usually there can be no compelling evidence that the implementation of the strategy contributed to attainment of the goal

The strategies are regarded often times as an instrument for the betterment of inter-institutional collaboration, and lessas an accurate expression of a projected policy. This view has some rationales: (i) the strategy is formulated in the continuous flux of policy process and as such it is a collection of ongoing programs and activities and new directions of action; (ii) the centers of decision are located in several areas of the policy arena and the views of the initiators of the strategy might not converge with the perspectives of the decision-makers from these alternative decision-making centers.

However, it may be beneficial that they are assessed both in respect of the implementation and also the attained results.

Discussion

The paper argues that policy evaluation should be regarded as a distinctive area of the evaluation field. It looks into some of the most respected recent literature on evaluation and social policy and invites reflection regarding the concepts from the theory of program evaluation that could be useful in policy evaluation and what would be the limits and opportunities for the use of such concept. Concepts such as the theory of intervention, pre-evaluation, policy learning, meta-analyses or reflexive controls show a good promise of becoming applicable in the area of policy evaluation as well.

Some potentially useful approaches for the evaluation of policies are proposed: capitalizing on cumulative knowledge, the comparative analyses and reflexive controls. An alternative strategy would be to identify the "strategic variables" of a policy and assess the changes in its dynamic and the correlated variables that were foreseen to change as an effect of the variation of the former.

Although policy evaluation is a complex enterprise and the level of certainty of its findings is considerably lower than in program evaluation, it is a fertile ground for further inquiry. There are many directions of research where the existing knowledge is insufficient. For instance, what should the rules be for meta-analyses of evaluations of *different programs* addressing interrelated problems,which are implemented jointly in a wider frame of a policy? What are the limits and opportunities for prospective evaluations of policies with no real ownership?

Further analyses of the author will be dedicated to the integration of the concepts discussed here into an integrative model and to testing the usefulness of this modelfor the assessment ofpolicies, such as thoseplannedthrough strategy formulation.

Bibliography

Albaek, E. (1990). Policy Evaluation: Design and Utilization. *Knowledge in Society*, 2 (4), Winter: 6-19

Bickman, L. (1987). The Functions of Program Theory, in L. Bickman (ed.), *Using Program Theory in Evaluation*, New Directions for Program Evaluation, no. 33. San Francisco: Jossey-Bass

Bleahu, A. (2004). Oamenii vorbind despre ei. Participare şi evaluare socială - studiu de caz. *Jurnalul Practicilor Pozitive Comunitare*, V(3-4), 17-28

Bochel, H., Duncan, S. (2007). *Making Policy in Theory and Practice*. University of Bristol: The Policy Press

Cace, C. (2002). Evaluarea programelor sociale. *Jurnalul Practicilor Pozitive Comunitare*, III(3-4),14-32

Palumbo, D.J., Hallet, M. A. (1993). Conflict Versus Consensus Models in Policy Evaluation and Implementation. *Evaluation and Program Planning*, 16 (1), January-March, 38-52

Campbell, D. (1969*)*. Reforms as experiments. *American Psychologist* 24 (April), 409-429

Clarke, A., Dawson, R. (1999). *Evaluation research. An introduction to Principles, Methods and Practice*. Sage Publications

Colebatch, H., Hoppe, R., Noordegraaf, M. (2010). *Working for Policy*. Amsterdam:Amsterdam University Press

Cronbach, I.J. and Associates (1982). *Designing Evaluations of Educational and Social Programs*. San Francisco: Jossey-Bass

Donaldson, S.I, Scriven, M.,(Eds.) (2009). *Evaluating Social Programs and Problems. Visions for the New Millenium*. New York, London:Psychology Press

Fetterman, D.M. (2000). *Foundations of empowerment evaluation*. Thousand Oakes, CA: Sage

Gerston, L.N. (1997). *Public Policy Making: Process and Principles*. Armonk, NY: M. E. Sharpe

Greene J.G. (1994). Qualitative program evaluation: practice and promise, in N. K Denzin and Y. S. Lincoln (eds.). *Handbook of Qualitative Research*. Thousand Oaks, CA: Sage, 530-534

Hill, M. (2009). *The Public Policy Process*, Fifth Edition. Person Education Limited

Howlett, M., Ramesh, M. (2003)*Studying public policy. Policy cycles and Policy Subsistems*, 2nded. Oxford University Press

Hogwood, B. W., Gunn, L. A. (1984). *Policy Analyses for the Real World*. Oxford University Press

Jenkins, W.I. (1978). *Policy Analysis: A Political and Organizational Perspective*. New York: St. Martin's Press

Knoepfel, P., Larrue, C., Varone, F., Hill, M. (2007)*Public Policy Analyses*, University of Bristol: The Policy Press,

Leeuw, F. L. (1991). Policy Theories. Knowledge Utilization and Evaluation, *Knowledge and Policy*, 4 (3), Fall:19-37, 73-91

Măţăuan, G. (1999). *Evaluarea programelor sociale*. Bucureşti: FIMAN.

Mertens, D.M. (1999) Inclusive Evaluation: Implications of transformative theory for evaluation. *American Journal of Evaluation*, 20 (1): 1-14

Mosse, D., Farrrington, J., Rew, A. (2002). *Development as Process. Concepts and methods for working with complexity*, 4th ed., Routledge, London and New York: Taylor & Francis Group,

Neguţ, A., Nicolăescu, V., Preoteasa A. M., Cace, C. (2011). *Monitorizare şi evaluare în economia socială*. Bucureşti: Editura Expert

Parsons, W. (1995). *Public Policy. An introduction to the Theory and Practice of Policy Analyses*. Edward Elgar Publishing Limited

Preoteasa, A. M. (2004). *Evaluarea programelor de dezvoltare comunitară*, Jurnalul Practicilor Pozitive Comunitare, III(3-4), 52-57

Pressman, J.L., Wildavsky, A.B. (1984). *Implementation: How Great Expectations in Washington are Dashed in Oakland*, 3rd ed. Berkeley: University of California Press

Patton M.Q. (1986). *Utilization-Focused Evaluation,* 2nd ed,. Newbury Park, CA: Sage

Palumbo, D.J., Hallet, M.A. (1993). Conflict Versus Consensus Models in Policy Evaluation and Implementation. *Evaluation and Program Planning*, 16 (1), January-March:38-52

Rew, A., Brustinow, A. (2002). The resolution and validation of policy reform: illustrations from Indian forestry and Russian land privatization, in Mosse, D., Farrrington, J., Rew, A. (ed.), *Development as*

Process. Concepts and methods for working with complexity, 4th ed. . London and New York:Routledge, Taylor & Francis Group

Rist, R.C. (1995). *Policy evaluation*, The International Library of Comparative Public Policy. England:Edward Elgar Publishing Limited

Rose, R. (1993). *Lesson-drawing in Public Policy. A guide to learning across time and space*. New Jersey:Clatham House Publishers, Inc. Clatham,

Rossi, P.H., Freeman, H. E. (1993). *Evaluation: A Systematic Approach*, 6th ed, Beverly Hills, CA: Sage

Shadish, W.R, Cook, T. D. and Leviton, L. C. (1991). *Foundations of Program Evaluation: Theories of Practice*. Newbury Park, CA: Sage

Strasser, S., Deniston, O. L. (1978). *Pre-And Post-Planned Evaluation: Which Is Preferable?*. Pergamon Press

Stufflebeam, D. L. (2001). *Evaluation Models*, New Directions for Evaluation, no. 89. San Francisco:Jossey-Bass

Weiss, C.H. (1986). The Many Meanings of Research Utilization, M. Bulner et al (eds), Social Science and Social Policy. London: Allen and Unwin, 31-40

Weiss, C.H. (1998). *Evaluation*. Prentice-Hall, UK

Walker, R., Duncan, S. (2007). Policy Evaluation in *Making Policy in Theory and Practice*. University of Bristol: The Policy Press, 169-190

Wollmann, H. (2007). Policy Evaluation and Evaluation Research, in Fischer, F., Miller, G. J., Sidney, M. S., *Handbook of Public Policy Analyses. Theory, Politics and Methods*. CRC Press, Taylor & Francis Group, 393-402

Zamfir, C. (1977). *Strategii ale dezvoltării sociale*, Bucureşti: EdituraPolitică

Görgens, M., Kusek, J. Z. (2009). *Making Monitoring and Evaluation System Work. A Capacity Development Kit*, The International Bank for Recontruction and Development /The World Bank, available at https://openknowledge. worldbank. org/handle/10986/2702

Zamfir, C., Stoica, L., Stănculescu, M. S. (2007). *Dezvoltare socială. Capacitatea instituţiilor. Ghid de evaluare*. Bucuresti:Institutul de Cercetare a Calităţii Vieţii, Academia Română, Seria Dezvoltare Socială

Annexes

Table 1. Types of evaluation

Type of evaluations	When to use	What is shows	Why it is useful
A: Formative Evaluation "Evaluability" Assessment Needs Assessment	1. During the development of a new program 2. When an existing program is being modified or is being used in a new setting or with a new population	3. Whether the proposed program elements are likely to be needed, understood, and accepted by the target population 4. The extent to which an evaluation is possible, given the goals and objectives of the evaluation and the program	5. It allows modifications to be made to the plan before full implementation begins 6. Increases the likelihood that the program will succeed
B: Process Evaluation Routine Monitoring	7. As soon as program implementation begins 8. During operation of an existing program	9. How well the program is working 10. The extent to which the program is being implemented as designed 11. Whether the program is accessible and acceptable to its target population	12. Provides early warning of any problems that may occur 13. Allows programs to monitor how well their program plans and activities are working
C: Outcome Evaluation Objectives-Based Evaluation	• After the program has made contact with at least one person or group in the target population	• The degree to which the program is having an effect on the target population's behavior	• Tells whether the program is being effective in meeting its objectives
D: Economic Evaluation Cost Analysis, Cost-Benefit Analysis, Cost-Effectiveness Evaluation, Cost-Utility Analysis	14. At the planning stage, using cost estimates/ projections 15. During operation of a program, using actual costs	• The resources that are being used in a program and their costs (direct and indirect) compared to outcomes	• Provides program managers and funders with a way to assess effects relative to costs
E: Impact Evaluation	16. During the operation of an existing program at appropriate intervals 17. At the end of a program	• The degree to which the program meets its ultimate goal	• Provides evidence for use in policy, funding, and future programming decisions

Source: Centers for Disease Control and Prevention, 2006, apud Görgens, Kusek, 2009

Table 2. Evaluation types and the time perspective

Time perspective	Evaluation question	Illustrative evaluation method	Counterpart formative evaluation question	Illustrative evaluation approaches
Extensive past	What worked?	Meta-analyses Systematic review	How did it work?	Systematic review
Past	Did the policy work?	Retrospective evaluation	How did it work? /not work?	Retrospective interviews Participative judgment (connoisseurship studies) Retrospective case study
Present	Is the policy working?	Monitoring Interrupted time series Natural experiments	How is it working? /not working?	Process studies Implementation evaluation Ethnography
Present to future	Is there a problem?	Basic research Policy analyses	What is the problem?	Basic research Rapid reconnaissance
Close future	Can we make this policy work?	Prototypes Micro-simulation	How can we make this policy work?	Theory of change Participative research Action research
Future	Will this policy work?	Program evaluation (impact or summative evaluation) Random assignment Matched designs Cohort designs Statistical controls	How will it work? /not work?	Retrospective Laboratory evaluation
Expansive future	What policy would work?	Prospective evaluation Micro-simulation Laboratory experimentation Gaming	How would it work?	Laboratory evaluation Delphi consultation Gaming

Source: Bochel and Sue Duncan, 2001: 175