Policy instruments, public deliberation and evaluation processes

POLICY INSTRUMENTS,
PUBLIC DELIBERATION
AND EVALUATION PROCESSES

Joan Subirats

Joan Subirats is Professor of Political Science and Administration at the Universidad Autónoma of Barcelona. He gave a seminar on "Public Policies in Spain: An Overview" at the *Center for Advanced Study in the Social Sciences* of the Juan March Institute on May 25, 1993.

In the beginning God created the heaven and the earth.

And God saw everything that He made. "Behold," God said, "it is very good".

And the evening and the morning were the sixth day.

And on the seventh day God rested from all his work. His archangel came then unto Him asking, "God, how do you know that what you have created is 'very good'? What are your criteria? On what data have you based your judgement? Aren't you a little too close to the situation to make a fair and unbiased evaluation?" God thought about these questions all that day and His rest was greatly disturbed. On the eighth day God said: "Lucifer, go to hell".

Thus was evaluation born in a blaze of glory. (Patton 1986, 1.)

Introduction

Evaluation can not only serve as a means of measuring or appraising the performance of policy instruments but can also foster consensus building among the different actors involved. This is done through a social learning process that involves discussing the criteria used to evaluate how effective the measures adopted are in terms of the problem or problems a specific policy is intended to attack. In this latter sense, evaluation is also a policy instrument, particularly in the case of policies that require a change in human behaviour or where there is little consensus about which variables should be used to measure the effectiveness of the policy adopted.

In this paper I will not attempt to discuss the characteristic features of evaluation studies so much as the evaluation process itself, the way the various actors contribute to the evaluation, how this process relates to social learning systems and how the outcome of the evaluation can subsequently be used to reformulate policies and the way they are implemented. I will mainly discuss environmental policies as this is a field in which these ideas are particularly applicable.

My working hypothesis is that institutional and cultural factors have a major influence on which policy instruments are selected, how they are applied and how they perform. It strikes me as unreasonable to maintain that the choice of any given policy instrument and the evaluation of its effectiveness should be based solely on technical considerations. The choice between economic or administrative measures in order to control pollution will be based on

ideological beliefs and philosophical tradition as much as on strictly technical reasons. To put this very simply: people who are in favor of regulating the market in order to control resources such as air and water, heretofore scarcely marketed commodities, are people with a favorable opinion of regulatory devices, while people who are reluctant to let market considerations influence any aspect of society will oppose such regulations.

Nor does it seem possible to maintain a clear distinction between objectives and instruments when the constant interaction between the two is obvious. It has been said that "demonstrating that there is a problem which can he attacked by one's favourite instrument is a very real preoccupation of participants in the policy process" (Majone 1989, 117). Once the truth of this is acknowledged, it can safely be said that the performance of the instruments selected (and therefore their evaluation) will depend not only on their theoretical technical effectiveness, but also on the political, social and governmental environments in which they operate.

In this context evaluation is a weapon that all participants can use to reinforce their favorite points of view, but this apparent weakness (partiality) can become a strength, a way of building consensus or of charting progress, if evaluation is made from what might be termed a "multiple" or "pluralist" prospective.

1. On Evaluation

Although there is some confusion surrounding the term, "evaluation" can be said to be traditionally linked to analyzing the results of specific policies or programs and an essential part of the policy process. It has been said that "evaluation research" makes it possible to relate objectives to results in a perspective not unlike that of productivity analysis (Dente 1989, 75-76; Morehouse 1972). Seen from this perspective, evaluation should involve:

    (a) enumerating the program's general aims and specific objectives;
    (b) discovering measurable indicators of these objectives;
    (c) collecting data about indicators for those people or things affected by the program and for an equivalent non affected control group; and

(d) analyzing the data on what has been done and the controls introduced in the light of the program's initial aims and objectives (Weiss 1972).

Taking these as their point of departure, a number of writers began attempting to give the evaluation process a scientific basis by introducing experimental or quasi-experimental techniques (Campbell and Stanley 1966; Langbein, L.I. 1980; Rossi and Freeman 1985). Critics of this approach have pointed out not only the difficulties involved in fully developing its initial hypotheses but also how little it reveals once it has demonstrated the difference between what was intended and what was actually done. These exercises in evaluation are somehow limited to viewing policy on its own terms. The value of what has been done is judged according to standards inherent to the policy itself, which implies accepting the policy goals and ignoring other, external values which might question the "rightness" of this policy (Jones 1992, 242).

(a) The limitations of the classical approach

It is a well-known fact that the classic approach to evaluation requires a series of conditions which frequently render it impossible or make its final conclusions scarcely relevant. It assumes that the decision-maker is an individual and acts as such. It assumes that his desires are consistent and that his preferences can therefore be assigned an order and the likelihood of certain things occurring evaluated. It assumes that the main thing is the outcome of the decision, not the process leading up to it. Intermediary or consensus-building factors are not part of the hardcore analysis. "Decisionists look upon policy problems as if they were puzzles for which, given clear goals and sufficient information, correct solutions always exist and can be found by calculation rather than by the exercise of political skills" (Majone 1989, 19-20).

The classical approach is somehow based on the assumption that it is possible to give scientific (objective) responses to the questions that comprise the evaluation process. However, the very experience of evaluations with the greatest scientific pretensions reveals how the experimental ingredient in evaluation *programs* was gradually reduced or watered down until it reached the point where it became evident that "social reality differs substantially from

physical reality" (House 1982; along the same lines: Monnier, 1987). The institutional framework, the set of actors and the interests at stake all varied according to the scenario, but intermediation and political commitment consistently outweighed pure program design. This indicates that people who stress the usefulness of evaluation, its ability to change behaviour and processes, or those who are seeking answers about values rather than just information about whether a given policy has met its objectives or not, would do well to start looking for a more plural (more "social") alternative to the necessary evaluation process.

(b) Policy and evaluation

Public policy is inevitably the subject of political dispute and not just in terms of values ("is this fiscal policy really fair?"), but also in terms of the instruments used to implement these policies ("are these really the most appropriate economic measures for increasing industrial production?") (Jones 1992, 25). It is difficult to keep evaluations strictly separate from, and uncontaminated by, politics. There are a number of reasons for this: evaluations are inevitably part of the atmosphere of political decision-making that surrounds the program being evaluated; moreover, the very content of the evaluation, its final value judgement, has a clearly political component; and lastly, evaluation can at worst be no more than a political instrument to support or criticize a particular program (Palumbo 1987, 12). The problem is not so much how to reduce the invariably political content that is present to a greater or lesser extent in all evaluations, as how to avoid reducing the evaluation process to a mere ritual designed to ratify or attack a particular policy.

The definition of politics as "something that cigar-smoking males do behind closed doors" (Palumbo 1987, 17) is unacceptable, just as it is unacceptable to confuse politics with inter-party competition. It is in the interaction between actors and their resources (power, influence, authority) that the basis of the intermediation inherent to all politics is to be found. It is precisely the recognition of this plural basis that defines our approach to the evaluation process, an approach in which the inclusion of these multiple decision-makers and interests is a determinative factor.

The intimate relationship between the political scenario and the evaluation process is illustrated by the fact that every program or action to be evaluated has its political decision-makers, its sponsors, the administrative personnel who have been involved in it, staff, clientele and interest groups who have in some way been affected by it. Evaluation is produced by and related to politics and hence becomes part of the political decision-making process. Above all, however, evaluation is somehow "political" in that it makes a certain value judgement which, though attempting to be objective, nevertheless involves taking a particular stance (Weiss 1987). One of the points I will examine in this paper is how to avoid turning evaluation into a mere "advocacy game" without necessarily accepting the legendary attributes of the scientific approach.

(c) Pluralist evaluation

A different concept of evaluation has been taking shape for some time now and it is one that does not attempt to evade the inevitable political aspects of evaluation or to neglect the need to acknowledge the social interaction inherent in all public action programs. Instead it attempts to combine the presence of the principal actors in the evaluation discussion with the necessary subsequent use of their findings. Although this kind of evaluation has not as yet been given a definite label, "pluralist evaluation" would appear to describe what other authors refer to as "naturalistic evaluation", "stakeholder evaluation", "utilization-focused evaluation" or "multiple evaluation" (Guba and Lincoln, 1981; Bryk 1983; Patton 1986; Majone 1989).

All these approaches share certain scepticism about the possible usefulness of evaluations that claim to be partial and objective. The pluralist concept maintains that it is not enough to base evaluation on purely empirical information without taking personal values and opinions into consideration. "In fact, values and opinions count a great deal in evaluation, not only because of the ambiguity of the outcomes in practice (...) but even more because of inescapable disagreements about the kind of evaluative criteria that are meaningful, fair or politically acceptable in a given situation" (Majone 1989, 168). These ambiguities or problems in focus cannot be resolved simply by employing more and better measurement techniques, but

also require a great deal of public discussion and persuasion. Analytical instruments, with their varying degrees of objectivity, would make more sense and become more operative were evaluation standards first defined in this public deliberation process.

The truth is that the actors involved in any public action program are always quick to offer their critical opinions. In this sense any evaluation made by specialists is nothing but just another opinion. Policies and policy instruments used in public action programs are constantly appraised and evaluated in the light of the particular logic and outlook of the people directly or indirectly involved. These multiple approaches are the very core of the policy design process in a pluralist society.

2. Evaluation as a Policy Instrument. Evaluation as a Social Learning Process

This paper will attempt to defend an approach to evaluation which does not simply emphasize the extent to which the internal objectives of the policy itself are or are not met, but which acknowledges the discussion about the basic values of the policy itself, the analytical process and the findings obtained, and aims to involve the principal actors in the evaluation process.

Evaluation is thus understood to be one more example of social pluralism and an admission that a number of different value systems exist. Viewed from this perspective, evaluation becomes part of the process whereby the members of society learn how to solve collective problems (Stake 1975; Monnier 1987). Examining the program to be evaluated becomes less a matter of discovering the "truth" about the effects of a particular program and more a way of revealing its different "truths" in an attempt to strike a balance which will be greater than the simple sum of the opinions of the individuals involved. As has recently been stated, "possibly the most effective single way to reduce impairment is to get into circulation a greater variety of messages, (a) competition of ideas" (Lindblom 1990, 293).

Emphasizing evaluation as an instrument for shaping confidence and creating consensus means involving everyone affected by the program or policy to be evaluated in an

attempt to gradually achieve the desired "objectivity". This entails getting both specialists and members of the general public to jointly discuss and explore the ramifications of the program or policy rather than simply providing a summary of some scientific premises that are always questionable. This would improve the policy being advocated, submitting it to a specific form of "multiple advocacy" (Majone 1989, 40).

The evaluator would not act alone, arbitrarily deciding whether the program under discussion is good or bad. Instead, he would act as a mediator between the different opinions. His scientific and technical background would not make him an objective and impartial judge, but would enable him to act as a negotiator, evaluating the need for information, answering questions, explaining consequences and attempting to reach an agreement about criteria and priorities (Ballart 1992).

(a) The limitations of an exclusively technical or professional approach and the social learning process

It is not my intent to question the importance of professionals in any field nor the significant role that evaluators can and do play in any process aimed at determining the usefulness of a particular policy. It is obvious that our society is increasingly dependent on specialists in a variety of fields. However as our dependence on specialists grows so too does our scepticism about their true capabilities and knowledge. We are sceptical because we have seen how the freedom of action of individual specialists or professionals is sometimes compromised. We are sceptical because we have seen certain theories or provisions backfire or because these same specialists have failed to foresee their effects. As if that were not enough, we are constantly hearing about partly or wholly contradictory theories and studies, all presented by the highly qualified individuals or teams. As a result, certain sectors of the public have joined consumer and other organizations in order to protect themselves from the corporatism of many professional associations. Moreover, there has been a considerable increase in the number of clients, patients, customers and other members of the public who take legal action against specialists.

This may seem incongruous in a society where specialization is becoming the general norm. Science and its technological applications have developed to a degree unimaginable just a few years ago. Today the welfare and economic development of any country are commonly gauged by the level of its "science and technology". Despite this, there are grave doubts about using all this scientific potential to resolve increasingly complex social problems. These doubts which stem from the very errors, subordinations, weaknesses and inconsistencies of scientists and their solutions, are aggravated by a democratic system of collective decision-making that is difficult to reconcile with technocratic logic or ideas that bring to mind a scientifically guided society.

It has been observed that nowadays people are more aware of reality, they have more information and technological expertise and all this makes decision-making more complex. As reality becomes more complex and chaotic, professional skills become increasingly relative (Schön 1983). This makes it necessary to know how to deal with indeterminate positions and conflicts of values. Complexity, instability, uncertainty and conflicts in value systems cannot be avoided or solved by applying rational techniques that attempt to provide specific solutions to well-defined, standard problems. Some practitioners manage to handle these complex situations by applying a vague mixture of intuition, skills and knowledge that results in a sort of "knowing-in-action" that is often hard to explain but which is based on the process of acting, accumulating the experience of these actions and developing a special sensitivity that enables them to detect the "winning habits" in a "reflection-in-action" process that can be synthesized as professional expertise (Schön 1983, 21). These mixtures of intuitive judgement and knowledge of the subject in question have been described as "non-logical processes" (Barnard), "the art of judgement" (Vickers) and "tacit knowing" (Polanyi).

This sort of insight, hitherto related to basically individual actions, could perfectly well be applied to organizations, institutions and groups such as policy communities and policy networks, which though less structured are no less tangible. Here too it should be possible to learn from experience, from the interaction between the different members or factions of a particular group. If evaluation practices were given a pluralist focus and the various actors affected by the policy in question encouraged to express their opinions, these practices would truly become instruments of social learning. Were resources like knowledge and trust

(Klok 1992) to be shared, both belief and disbelief would be expressed, winning habits shaped and situations from which something could be learned created, thereby giving the evaluation process its fullest meaning as a policy instrument.

Evaluation could thus become a formal occasion for examining the way specific programs of action have been experienced by the people affected and for discovering what they believe to be the successes and failures of these programs. In a situation where pluralism is accepted and there is a real desire to improve the operating environment, evaluation is a summing-up of the social learning resulting from the interrelations of the policy community. Policy errors will not be condemned, but will be viewed as sources of information and clues to subsequent actions to be taken.

> "In this model, citizens, functionaries, social scientists and other experts do what they have learned and they learn what they have done" (Lindblom 1990, 219).

(b) Citizens as analysts, analysts as citizens

When a particular policy is classified as a success or failure, this frequently means that it has been viewed from a narrow managerial focus, more concerned with meeting internal policy goals or exercising effective administrative control than with the program's ability to respond to the needs of the various individuals and groups affected. It also means that valuable information for political decision-makers is being ignored. Moreover, efficiency and effectiveness, which are typical values of evaluation studies, often clash with the attitudes of policy implementers who are more concerned with the professional skills and decision-making powers of the specialists who designed the policy.

Many public policies currently being implemented are based on the assumption that they can and will change the public's style of operations or their scale of preferences. In some cases public policies are expressly aimed at certain groups of citizens while others affect the public as a whole. In these cases, encouraging public participation not only expresses a general desire to make life more democratic, but is absolutely essential if policies are to meet with greater success.

Deciding who should participate and how naturally poses problems. The people implicated in a particular policy (politicians, the experts or officials who implement the policy, special interest groups, individual citizens, outside experts...) are the people who should be involved in any pluralistic evaluation process. The process will then strike a balance between standard professional evaluations, with their almost exclusive emphasis on expert opinions, and the criticism that any action taken by the public powers provokes in one sector of society or another. All the individuals or groups affected have different opinions about the evaluation criteria to be used and the emphasis to be placed on particular aspects of the policy process. General standards of performance, such as legality, legitimacy, economy, effectiveness, efficiency and the ability to provide responses to social needs are the specific concern of certain of the actors involved (politicians, judges, consumers...). It should also be acknowledged, however, that each of these criteria can be more easily applied to a particular part of the policy process than to the process as a whole (for example, legality is more aptly linked to process, economy to inputs, etc.) (Majone 1991).

But these multiple criteria mirror reality and must form part of the evaluation process, particularly if the aim is to subsequently act on the conclusions drawn. These are not so much the result of agreement as to what has been achieved, but of how much the policy community has learned about the problems revealed and the mistakes made.

The general criteria Wildavsky listed years ago (Wildavsky 1979, 253) still apply. The policy community must understand what is in it for them, recognize the differences between small and large changes, and be involved continuously so that they can learn from experience (Wildavsky 1979, 253). The ability to gradually shape a policy community that is sufficiently open yet specialized, that recognizes disappointment as a constant policy test, is the best way to insure improved policy performance. Evaluation, as an opportunity to discuss objectives and results, may well be the most suitable instrument for doing this.

(c) Using evaluation

All this is easier said than done, but if it can be achieved then evaluation will certainly be used to improve the program involved. Criticisms of the evaluation process generally point to

> *(i)* the weakness of the methodology used in the evaluation. Often the procedure employed in evaluating a particular program causes the validity and credibility of the evaluation itself to be called into question.
>
> *(ii)* irrelevance. The outcome of an evaluation may be inadequate at a particular moment or may simply fail to provide the decision-makers with any new information.
>
> *(iii)* infra-utilization. The information obtained is either not made public or not subsequently acted upon.

The idea of evaluation continues to be intrinsically linked to its ability to reveal whether or not a particular policy has had the desired results. As mentioned earlier, evaluation tends to be based on the decision-makers' own terms and implies that they have decided how things should be done and subsequently taken actions designed to get results. Evaluation is intended to examine these results and indicate the extent to which the intended results have been achieved. It is assumed that evaluation is an objective, internally coherent process when in fact the evaluators' very objectivity and supposed impartiality can be the underlying cause of problems in getting results.

The general impression is that any short-term evaluation designed to classify a policy or program as a success or failure tends to conclude that the policy or program involved has had little effect. Because actions involving public powers are so complex, evaluation must be made less "peripheral" and viewed more as a learning tool than an objective measure of performance. For this reason it is absolutely essential that both the different levels of public powers involved in policy-making and the people who will be affected by these policies take part in the evaluation process.

There are so many different evaluation criteria, so many standards of accountability, that the analyst cannot be expected to judge something that belongs more to the field of politics than to the realm of analysis. Instead, his role is to provide bases for discussion that are as

for discussion that are as appropriate as possible. "The need today is less to develop 'objective' measures of outcomes -the traditional aim of evaluation research— than to facilitate a wide-ranging dialogue among advocates of different criteria" (Majone, 1989, 183).

Rather than limiting us to a few pre-defined analytical categories we could take advantage of the different viewpoints of the people involved, their different sources of information and their different experiences with particular policies in order to discover effects and connections, links between processes and outcomes and examine how different factors interact with time.

Focusing more on what people really do, the real implications of their problems, their basic situations and the roles they play; charting the way they interact, their strategies and arguments; accumulating and compiling their experiences; studying their basic assumptions and comparing facts with these arguments and assumptions, can all combine to form the basis for evaluation as a consensus-building instrument and provide a guarantee that it will be increasingly used as an instrument for improving policies (Hellstern 1985).

3. Pluralist Evaluation, Public Deliberation and Environmental Policy

(a) General features

It has been said (Bosso 1987) that environmental policy is much more "resistant" to becoming part of an accepted routine than are other policies and that environmental policy is therefore invariably more "political". This may in part explain why in some countries there is so little tradition of environmental policy (viz. southern Europe) and why it is often considered a "mobilizing" policy. This means that there is a bargaining arena and a structure of the agenda that we cannot consider fixed, with a fairly high level of conflict because of the disagreement over goals and casual theories as well as over procedural rules of the game. At the same time, the lack of consensus among the increasingly numerous experts in the field is gradually giving

the whole field of environmental policy a reputation for fluidity, openness to change and renovation. There is a lot of conflict between the different groups but the internal cohesion of the most active of them is on the rise. It is difficult to talk about stable and structured policy communities in countries like Italy (Lewansky 1990, 310) and Spain (Aguilar 1992).

Furthermore, institutional responsibility is highly fragmented, with various levels of government and the public administration being simultaneously involved (either actively or passively).

In Europe alone, the extent of environmental awareness differs greatly from one country to another. The EC and other supra-national agencies design policies and enact regulations that have little relation to the priorities on the national, regional or local agendas of certain countries. Numerous institutions split the responsibility for environmental issues and for developing specific regulations to control environmental deterioration.

These general features make environmental policy an appropriate field in which to experiment with pluralist approaches to evaluation that would make it possible to improve environmental action programs and control the application of specific regulations. As we all know, the actors most frequently involved in environmental policy (experts, managers, judges and jurists, economists, ecologists, politicians, the business community, the sectors of the community who are directly affected, the general public) often have radically different viewpoints.

Moreover, the general public, governments and experts are all increasingly concerned about the environment and this has made more information available, but it has also increased uncertainty about acceptable risk levels and shed doubt on forecasts about how certain variables are likely to develop (Hawkins and Thomas 1989). The abundance of available data and the proliferation of studies and forecasts have only increased uncertainty and highlighted the fact that the environment is not a linear system. Environmental policy has been labelled a policy of scarcity. This scarcity is sometimes artificially created by scientific or scientific and social consensus. The way society perceives scarcity is important. Equally important are

information and public awareness of the gravity of the current situation and the need to enact urgent measures to stem environmental deterioration.

Although greater social awareness of environmental problems has led to increased controls, it has done little to alter the conduct of the people who cause pollution. If pollution is to be controlled, then industries must be responsible for controlling their own operations and introducing environmental protection measures into their production processes. Social pressure and public concern over the environment can be vitally important in changing industrial attitudes.

All this would seem to indicate that evaluation can be effective as an instrument for consensus-building and as a means of improving environmental policies within a framework of public participation and social learning.

(b) Evaluating environmental policies

There is a great deal of doubt about the true effectiveness of recent regulations and how much environmental policies have actually done to change things.

All thoughts of evaluating environmental policy tend to be rooted in economic concerns or linked to civil engineering traditions. Cost-benefit analyses have demonstrated that political factors are crucial in determining costs and that this analysis technique was being used in a way that "looked neat, and seemed neutral to the naked eye. Reducing policy to numbers made people feel comfortable; the process gave a sense of clarity to issues that had heretofore defied easy definition" (Tolchin 1987, 279). In practice however, the more qualitative elements that could pinpoint benefits and define impacts are played down because they are considered too complicated to measure or because they are indivisible.

Moreover, the "Environmental Impact Evaluations" recommended by EC Directive 85/337 as a preliminary to any public action that will in any way affect the territory, are highly technical and use standards of measurement that are often questionable. Equally, they do not take into account the opinions of the people affected or those of the public in general, even though reference is made to "the aptness of the social environment" for the decision being

made (Gómez Orea 1988). "Public participation" takes place at the end of the analytical process and is merely a formality, similar to the "public participation" in city planning measures.

(c) Evaluation as an environmental policy instrument

Environmental policy is closely linked to regulatory measures enacted by specialized agencies connected with the different actors involved. Policy making and discretion are highly interdependent, particularly inasmuch as it is so difficult to define so-called risk situations in an objective or generally acceptable way (see Bosso 1987).

The interaction of policy makers, experts, bureaucrats, interest groups, users and the public should therefore be instrumental in measuring the extent of adherence to regulations and evaluating their effectiveness. The uncertainty surrounding risk situations and the limitations of science and technology recommend the use of public deliberation to determine what risks are acceptable and what kind of technological solutions, regulations or discretionary measures will be tolerated (Hawkins and Thomas 1989, 273).

Processes based on the techniques of "utilization focused evaluation" (Patton 1987, 122 ff.) or pluralist evaluation (Monnier 1987) can provide an adequate framework for intermediation processes that help confer legitimacy on the measures adopted. Indeed, environmental policies are often based on the relationship between a physical or biological reality that is felt to be deteriorating or endangered and what is scientifically desirable. All this is part of a process whereby the different political agents make the corrections they consider necessary from their ideological viewpoints or on the grounds of social consensus. Evaluation is presented as a chance to discover whether things have improved or not, but the definition of "improvement" will differ depending on the viewpoint involved. Natural science may furnish new information but this information will be subject to different interpretations. The more unanimous the scientific facts, the greater the possibilities of reaching an agreement. Thus there is a direct, though not always straight, line between scientific facts and social convictions. Here too, pluralist evaluation processes could help reinforce social convictions thereby increasing the chances that the measures adopted will be adhered to.

In terms of environmental policies, the relations between scientific certainty and social certainty are conclusive in gauging how successful certain measures (designed to change social comportment, practices that cause pollution, etc.) will be. Myriad social interests, the disparity of criteria applied or of positions adopted when judging whether the environmental policy measures enacted are correct or not, should increase the likelihood of social debate. The existence of different schools of thought and different parameters will not infrequently keep scientific discussion alive. (See the Asarco-EPA case in Scott 1988; see also Table 1, where position A is very infrequent in environmental problems, while position B allows a political decision that stops social debate, and position C offers room for a formal evaluation because of the social agreement on the non-controversial nature of the issue; position D, probably the most frequent situation in environmental policies, fits in very well to our pluralist approach of evaluation as an instrument.) Open, pluralist evaluation processes will quite possibly prove to be a suitable way to confer legitimacy on specific environmental policies. They will also serve as an instrument for social learning in which scientific facts will be given their just due and improvements or deterioration in the behaviour of the public concerned will be clearly revealed, thus shaping a common ground for attempting to improve the definition and implementation of environmental policies.

Although evaluation of environmental policies as described here should take account of the different opinions of the members of society, it should not underestimate the facts presented by scientists and other experts. By not ignoring the way society receives these facts and alters its behaviour to fit the circumstances, the evaluation process should even be able to improve on these scientific facts, serve as a social and organizational learning tool and be instrumental in improving the environment.

Table 1. *Scientific certainty and social certainty*

		Social Certainty	
		yes	no
Scientific Certainty	yes	A	B
	no	C	D

REFERENCES

Aguilar, S. 1992. *Políticas medioambientales y diseños institucionales en España y Alemania: La Comunidad Europea como escenario de negociación de una nueva área política.* PhD Thesis, Madrid (mimeo).

Ballart, X. 1992. *Evaluación de programas y servicios públicos.* Madrid: INAP.

Bosso, C.J. 1987. *Pesticide and Politics.* Pittsburgh: University of Pittsburgh Press.

Browne, A. and A. Wildavsky 1987. "What should evaluation mean to implementation" in Palumbo, ed., *The Politics of Program Evaluation.* Newbury Park: Sage, pp.46-172.

Bryk, A.S. 1983. *Stakeholder Based Evaluation.* San Francisco, CA: Jossey Bass.

Campbell, D.T. and J.C. Stanley 1966. *Experimental and Quasi-Experimental Designs for Research.* Skokie, Ill. Rand-McNally.

Cronbach, C.J. 1980. *Toward Reform of Program Evaluation.* San Francisco: Jossey Bass.

Dente, B. 1989. *Politiche Pubbliche e Pubblica Amministrazione.* Rimini: Maggioli.

Gómez Orea, D. 1988. "Evaluación de impacto ambiental" in *Ciudad y Territorio,* 75:5-32.

Guba, E.G. and Y.S. Lincoln 1981. *Effective Evaluation: Improving the Usefulness of Evaluation Results through responsive and naturalistic approaches.* San Francisco: Jossey Bass.

Hawkins, K. and J.M. Thomas 1989. "Making Policies in Regulatory Bureaucracies" in Hawkins and Thomas, ed., *Making Regulatory Policies.* Pittsburgh: University of Pittsburgh Press, pp.3-30.

Hellstern, G.M. 1985. "Assessing Evaluation Research." In Kaufmann, Majone, Ostrom, (eds.) *Guidance, Control and Evaluation in the Public Sector.* Berlin: De Gruyter, pp.279-312.

House, E. 1982. "Preface to Evaluation Studies". In *Evaluation Review Annual,* vol.7.

Jones, P. 1992. "Evaluation" in Harrop, M., ed., *Power and Policy in Liberal Democracies.* Cambridge: Cambridge University Press, pp.241-262.

Kelman, S. 1981. *What Price Incentives?.* Boston: Auburn House.

Klok, P.J. 1992. *A Classification of Instruments for Environmental Policy.* ESF (mimeo).

Langbein, L.I. 1980. *Discovering whether Programs Work: a guide to statistical methods for program evaluation.* Glenview, Ill: Scott-Foresman.

Lewansky, R. 1990. "La Politica Ambientale" in Dente, ed., *Le Politiche Pubbliche in Italia*. Bolonia: Il Mulino, pp.281-314.

Lindblom, C.E. 1990. *Inquiry and Change*. New Haven: Yale University Press.

Majone, G.D. 1989. *Evidence, Argument and Persuasion*. New Haven: Yale University Press.

_____ 1991. "Public Policy beyond the Headlines." *EUI Working Papers in Political and Social Sciences*. Florence: European University Institute.

Monnier, E. 1987. *Evaluation de l'action des pouvoirs publics*. Paris: Economica.

Morehouse, T.A. 1972. "Program Evaluation: Social Research versus Public Policy" in *Public Administration Review*, pp.868-874.

Palumbo, D.J. 1987. "Politics and Evaluation" in Palumbo, D.J. (ed.), *The Politics of Program Evaluation*. Newbury Park: Sage, pp.12-46.

Patton, M.Q. 1986. *Utilization-focused Evaluation*. Beverly Hills: Sage.

_____ 1987. "Evaluating Political Inherency. Practical Implications for Design and Use." In Palumbo, D.J. (ed.). *The Politics of Program Evaluation*. Newbury Park: Sage, pp.100-145.

Rossi, P.H. and H.E. Freeman 1985. *Evaluation: A Systematic Approach*. Beverly Hills, California: Sage.

Schön, D.A. 1983. *The Reflective Practitioner*. New York: Basic Books.

Scott, E. 1988. "Managing Environmental Risk: The Case of Asarco," *Case Program. Kennedy School of Government*. Harvard University: Cambridge.

Stake, R.E. 1975. *Evaluating the Arts in Education: A Responsive Approach*. Columbus: Merrill.

Tolchin, S.J. 1987. "The Political Uses of Evaluation Research: Cost-Benefit Analysis and the Cotton Dust Standard" in Palumbo, D.J., ed., *The Politics of Program Evaluation*. Newbury Park: Sage, pp.249-270.

Weiss, C. 1972. *Evaluation Research*. Englewood Cliffs, NJ: Prentice Hall.

Wildavsky, A. 1979. *The Art and Craft of Policy Analysis*. London: Macmillan.