Evaluation in public administration

Five logical approaches

STRUMENTI N. 3

STRUMENTI

Ufficio Valutazione Impatto
Impact Assessment Office

Senato della Repubblica

This paper was written by

BARBARA ROMANO, Turin University

ALBERTO MARTINI, Piemonte Orientale University

Evaluation in public administration

Five logical approaches

October, 2018

This short essay has the ambition of bringing some order to the confusion in the usage of the *term* evaluation. The classification we offer has five entries; (a) evaluation as a type of selection mechanism; (b) evaluation as a tool of controlling the performance of organizations; (c) evaluation as way of being accountable for what an organization has achieved; (d) evaluation as a tool for critically examine the implementation of a policy; (e) evaluation as assessment of the causal relationships between policies and outcomes—that is, for measuring the causal effect of a policy.

Sommario

Evaluation has long been called for in the Italian public administration and the term evaluation has a multifaceted and wide-ranging meaning, which changes depending on different contexts and situations: this sometimes leads to difficulties in communicating and in reciprocal comprehension between insiders themselves. This heterogeneity of meanings is also fuelled by the multidisciplinary nature of this professional practice, where subjects with very different disciplinary and cultural backgrounds coexist: sociologists, economists, statisticians, political analysts, psychologists, businessmen and town planners.

In view of the heterogeneity of knowledge-related needs and of the disciplinary stimuli that underlie it, we reckon that trying to find one single, all-encompassing definition for evaluation and its goals would be **futile**. On the contrary, we believe it is more useful to conceptually **organise** this heterogeneity, accepting it as a fact.

The suggested classification that follows is actually based on the attempt to *distil* the crucial fact-finding requirements and the way evaluation can meet them. Bear in mind two things: i) it's stylisation; ii) it is not an exhaustive classification, it simply represents the will to grasp the foremost aspects of this phenomenon.

From this standpoint, evaluation meets five types of concrete and specific fact-finding goals, which in practice can blend and get mixed up. Yet it is important to address them separately. Evaluation can therefore be a tool to:

(i) allocate resources *by identifying the most worthy recipients*;

(ii) *monitor performances* within organisations;

(iii) *account for the accomplishments* of a public action;

(iv) *critically analyse the implementation* of a project or of a public policy;

(v) *measure the effects* generated by a public policy.

Let's zero-in on each one of these fact-finding goals, highlighting their essential features and briefly outlining the analytical instruments the evaluator may resort to in order to put them to good use.

1. Finding worthy recipients

In this first field, the fact-finding goal of evaluation is that of providing a *comparative judgement of several subjects in order to identity the most worthy ones*. This goal reminds us of an almost primordial form of evaluation, the one we all face as children, when we are given the first marks at school. In public activity, we often witness the need to give marks to a number of subjects (people, actions, projects, institutions), usually in order to identify the best and, consequently, to better allocate scarce resources. However, sometimes it's just a matter of stimulating emulation-like behaviours by rewarding the best. Therefore, this type of evaluation is basically a way of formalising and standardising (bottom line: legitimising) the process that makes it possible to compare subjects and tell which one is better.

The concrete forms of evaluation that look to fact-finding in order to identify the best subjects range from the evaluation procedures (aka selection) of projects to the evaluation of personnel, to methods evaluating scientific research. What always underlies this type of evaluation is the need to *make a choice* when considering alternative options: the choice may be about public officials that should be given additional remuneration owing to a greater productivity; competing research projects that should be granted public funds; public works that are worth investing in.

Evaluation for identifying the most worthy subjects entails three specific steps:

- above all, it is necessary to set the *criteria*, namely the parameters for assessing the subjects: when there is more than one criterion, it is necessary to determine the *weight* of each criterion, as well as the point-assigning *metrics* based on each one;
- once criteria and weights have been established, the actual evaluation is conducted by *experts* that express an *opinion* – converting it into a score, based on the chosen metrics – of each subject, based on each criterion: that gives the score matrix, calculated as MxN (M being criteria, N being subjects);
- finally, the resulting judgements must be *aggregated* by adopting an explicit procedure, which may range from the simple weighted sum of the normalised scores, to the use of sophisticated multicriteria analysis techniques: aggregated judgements allow us to identify the best subjects among the evaluated ones, generating a ranking.

The concrete forms that the above-mentioned evaluation process takes on may vary from case to case. In particular, the content of the second step – where experts intervene – greatly depends on the nature of the analytical tools such experts use to formulate their judgement: this may range from the mere expression of subjective opinions legitimised by the professional prestige of the experts themselves and their specialised knowledge, to the use of sophisticated and complex analytical techniques, such as the cost-benefit analysis, which may be required, for example, when judging the value of alternative infrastructure projects.

2. Monitoring the performance of organisations

The fact-finding need underlying this second type of evaluation is fundamentally that of understanding *how well* a public organisation (or a part of it) is working. We usually talk about *performance*, meaning an organisation's set of relevant operating elements (such as costs, quality and overall activities).

This type of evaluation basically consists in a comparison of the *measurements* of what the organisation has produced (and how it has produced it) and the corresponding *desired values*, which represent what the organisation *should* produce (and how it should produce it). The term *indicator* is often used when referring to the performance's single aspects that must be assessed by comparing observed and desired values.

The desired values for each indicator can derive from sources that may vary considerably, depending on the specific requirements. In the simplest cases, they could be the values observed in the past for the same indicator, used as a comparison parameter to check whether there has been an improvement or a decline. Or, they can be *target values* set by the *management* to directly stimulate the different units or parts of the organisation and help them boost their performance. Or, they can be qualitative standards set by professional bodies or public authorities. Finally, they could be the values observed for the same indicators in similar organisations, therefore acting as *benchmarks*.

This kind of evaluation features five typical stages:

- the identification of the *dimensions* of the organisation's actions that are to be monitored;
- the definitions of the *indicators*, which characterise the organisation's performance in each dimension;
- the identification of the appropriate *desired (standard) values* for each indicator, namely the benchmarks that allow to distinguish between a good performance and poorer one;
- the systematic collection of *data* on the organisation's performance, to obtain the value of each indicator and compare it with the pertinent desired value;
- the interpretation of the observed performance's *difference* compared with the preset standards, in order to field *mending actions* aimed at improving inadequate performances.

In-depth analyses would be necessary to convince the reader that such a generic list of steps is not really superficial, but such in-depth studies lie beside the point of this brief paper, which is to highlight that the third step – *the identification of the desired (standard) values* – is the real challenge for those who want to conduct this type of evaluation.

On a critical note, evaluation procedures do not always confront such a challenge. In Italy, focus is – either directly or indirectly – on an evaluation of performances that generally tends to concentrate on the second step, *the definition of indicators*.

Examples of concrete activities that rely on this conceptual scheme, albeit being different, are *management control, quality certification* (now widespread in the public sphere) and many forms of *validation*. Even tools such *user satisfaction* surveys can fall within this type of evaluation: in this case, it is the users themselves that directly compare observed values and desired values by comparing the service they received (and the perceived quality) to their expectations, which in this particular situation act as the standard values.

3. Accounting for accomplishments

This third type of evaluation is underpinned by the need to describe and measure the *accomplishments* of specific forms of public action, in order to *account to* the most relevant *stakeholders*. This form of evaluation differs from the previous one, for at least three reasons.

Firstly, the *performance* analysis focuses on the gap between the performance and the preset standards: the idea is that of realising whether, and to what extent, the organisation is working well, or how distant it is from a desirable performance. Evaluation as a *tool for giving account* does not rely much on comparison; it rather highlights the *descriptive and communicative* sphere. Attention is prevalently given to aggregate dimensions (which offer an overall picture of the effort), and not to specific measurements and details.

Secondly, performance analysis is mainly used to measure and assess the performances of the organisation's single units. Evaluation as a tool for giving account, on the contrary, does not focus on the single unit, but on complex and multifunctional organisations, or on major public intervention programmes.

Finally, the most relevant difference is about the possible consequence of the two types of evaluation. Performance analysis results may be used to take managerial decisions, hence intervening directly on the organisational realities that show poor or inadequate performances: the information provided by evaluation as a way to give account has less impact on decision-making, as it carries a communicative value. Indeed, the purpose of such information is that of giving *third parties* an overall idea of the action fielded by the administration, of the reasons underlying such strategies, of the activities conducted during the analysed period and of the results obtained at the end of that period.

The key terms in this case are *transparency*, *accountability* and *legitimisation*: through this form of communication, public administration offices try to be more transparent and account to their stakeholders for the choices made and the strategies adopted, in order to recover, or consolidate, their legitimisation to operate.

The most relevant examples that fall within this category are, on the one hand, the complex *monitoring and evaluation* process that accompanies the granting of European Structural Funds; on the other, the vast movement in favour of a greater *accountability* of the Italian public administration, particularly towards citizens, through instruments such as the *term accountability report* and the *social accountability report*.

It may seem piddling to describe the evaluation of Structural Funds as a tool to account for accomplishments. We are aware that such an evaluation system pursues far greater fact-finding goals, which partially overlap the goals we assign to the other two types (hereunder described).

On the one hand, the evaluation of Structural Funds aims at identifying the criticalities of measure-implementation processes, in order to correct them while they are underway; on the other, it often refers to the need for evaluating the effects, the contribution, the added value or the consequence of the measures co-financed by the Funds.

We shall simply say that the peculiarity (and usefulness) of what we have defined "evaluation as a tool to give account" is, above all, *briefly illustrating* (with tables, graphs or just words) to one's financiers, or to one's contributors/users/electors, what has become of the resources made available to a Public Authority, to a Mayor, or to a Regional Cabinet. However, such illustration does not expect to actually identify criticalities or effects.

4. Analysing the implementation of a policy from a critical standpoint

This fact-finding need stems from the awareness that the enforcement processes of policies (and of any other public action) are unpredictable and can hardly be controlled from above: the high hopes put into public policies often become disappointment, and sometimes total failures.

Each public policy must shift from being an abstract idea to an administrative practice, from being a good intention to something that can be implemented through the actual use of the available resources, through the provision of real services, and through the proper enforcement of rules or prohibitions. An English expression efficiently describes this phase as a shift *from policy fiction to policy facts,* where *fiction* is the abstract idea of the policy that *policy-makers* have in mind (whether they are lawmakers, government members or chiefs of public administration offices). In order to become *facts*, these abstract ideas must go with the will, the interests and the resources of all the players that must transpose the guidelines and goals of public policies in their daily work (within and outside of the public administration sphere).

The fate of a public policy is often that of remaining a *dead letter* or of being implemented in a way that is very different from the one originally envisaged. This is a phenomenon known as *implementation deficit*, which can be witnessed at all latitudes and in all political systems. A vast literature (available especially in English) studies, documents and explains the difficulties that hinder the path towards the full implementation of a public policy. This literature has a number of names, the most common probably being *implementation research*.

When an evaluator wants to critically analyse the implementation of a policy or a project, s/he reconstructs – mainly through qualitative investigation tools – the policy-enforcing processes, in order to understand if, and why, they differ from the original idea. The prevalent fact-finding requirement in this case is not that of judging the performance, nor accounting for what has been done, nor measuring effects and consequences: the goal is actually that of *entering the processes* adopted to concretely enforce a policy, talking to the players involved in these processes and trying to shed light on critical aspects, inconsistencies and problems that hinder the path to implementation.

The typical elements of this kind of evaluation are the following:

- describing the concrete implementation of the policy, the timing, the main stages, the achievements and the players truly involved;
- pointing out whether the original intentions of the policymaker were modified as the implementation process was underway;
- investigating the plausible reasons for such differences and how these differences affected (if they did) subsequent decisions;
- analysing specific aspects such as the organisational context in which the policy was implemented, the methods adopted for the selection and treatment of the policy recipients, the way the officials handled the single operating activities.

Many different investigation tools can be used to achieve these fact-finding goals. Yet they will tend to focus on quality (hence aimed at gathering opinions, perceptions, behaviours, motivations among the implementation process players), rather than quantifying indicators or measuring activities. Social sciences field a vast array of qualified investigation techniques that can help achieve this goal: above all, interviews (structured as questionnaires, semi-structured, conversational) including focus groups, aka collective interviews, brainstorming techniques, direct observation. But all of these are just techniques whose success depends on the way the evaluator uses them, on whether s/he has the due insight and is able to *ask the right people the right questions*.

5. Measuring policy effects

The fact-finding purpose of this type of evaluation is that of understanding *what works*, namely, what public policy instruments efficiently lead to the desired change. The preliminary steps for correctly setting the evaluation of a policy's effects are the following: defining *on what* the policy has an effect and *what* is causing the effect.

It is therefore necessary to identify the problematic facets to be modified: the behaviours/conditions that represent the problem the policy intends to address are to be converted into *observable and measurable* elements. It's a matter of identifying the *outcome variables*. Avoid making a common mistake: identifying an outcome variable is different from estimating the policy's effect on that variable. Whether such an effect comes about or not (its extent, if it's statistically relevant) is a totally different issue. It is easy to get mixed up, even because of the existing literature, which defines the aforesaid outcome variables as "choice of indicators". It is obviously also possible to use the term *indicator*, provided that one accepts that the *indicator, per se*, does not indicate the policy's effect: sometimes this is the meaning given to indicators, which are labelled – erroneously, in our opinion – as "impact indicators".

Once established the outcome variables in which the policy effects are to be sought, it is still necessary to identify the *treatment* represented by the public policy that is to be evaluated. That means it is necessary to exactly establish *what* is causing the effect. One is immediately tempted to say that what is being sought is the public policy's effect, but that alone is clearly not enough. Other than for a few exceptions, public policies are *a mix of measures of varying natures, implemented through different timings and intended for different recipients*. Thus, it is necessary to simplify things and identify an outcome variable that, in its simplest form, can be a dichotomic variable, representing the absence/presence of the treatment, or better still, the exposure or non-exposure of the population individuals to the treatment.

Once highlighted *on what* the effects are sought and *what* causes the effects, it is necessary to devise a method for plausibly estimating such effects. The fundamental methodological problem that this type of evaluation must tackle is the following: in order to establish the effect produced by a given measure, we must somehow reconstruct what would have happened without such measure being implemented, namely, the counter-factual situation.

A policy's effect is *defined* as the *difference* between two values of the outcome variable. However:

- one of the two is *observable*, among the subjects exposed to the policy, *after* the exposure (factual value);
- the other value is *hypothetical*, meaning that it represents what *could have been* observed among the same subjects had they not be exposed to the policy (counter-factual value). It is worthwhile reasserting that such difference cannot be calculated directly, because while the first of the two terms of difference can be directly observed (in line of principle, at least), the term *can never be* observed, since it's the very implementation of the policy that makes it impossible to observe. Strategies for evaluating the effects of a policy are therefore essentially strategies that help give an approximate, *credible* counter-factual value, using all the information available and the appropriate statistical techniques. The more the strategy adopted for estimating the counter-factual value is credible, the more the evaluation of the effects will be plausible.

From this standpoint, the evaluation of effects does not produce unshakeable certainties, but it can lead to *reasoning,* which can enrich the public decision-making process. One should be aware that using figures and statistics, even in a meticulous way, does not necessarily produce unquestionable results; it simply makes available – to policymakers and to the community – solid and credible reasoning in favour or against a specific policy strategy. Effect analysis is seldom integrated by the cost-effectiveness analysis, which estimates the cost needed to obtain a *unit* of effect: for example, how much it costs to have one extra student enrolled in university *owing* to a tertiary education support programme, or how much it costs to have one extra working employee *owing* to a policy aimed at stabilising precarious workers. Policymakers should use the results of the cost-effectiveness analysis to compare the different policy options that address the same social issue.

Bibliography

Lennon M., Corbett T. a cura di (2003) *Policy into Action*, The Urban Institute Press, Washington.

Martini A., Sisti M. (2009), *Valutare il successo delle politiche pubbliche*, Il Mulino. Bologna

Pressman J., Wildavsky A. (1973), *Implementation - How great expectations in Washington are dashed in Oakland,* University of California Press, Berkeley, CA.

SENATO DELLA REPUBBLICA

UFFICIO VALUTAZIONE DI IMPATTO
IMPACT ASSESSMENT OFFICE

www.senato.it/ufficiovalutazioneimpatto