METAEVAlUATION AS A TOOL FOR THE IMPROVEMENT AND DEVELOPMENT OF THE EVALUATION FUNCTION IN PUBLIC ADMINISTRATIONS

María Bustelo (Associate Professor. Universidad Complutense de Madrid)
Departamento de Ciencia Política y de la Administración II
Facultad de Ciencias Políticas y Sociología
Universidad Complutense
Campus de Somosaguas s/n. 28023 MADRID
mbustelo@cps.ucm.es

Nowadays, in most European countries, evaluations are becoming a reality, so that there is a certain increase on empirical references of evaluation experiences in different policy fields. However, most of these experiences have not been already studied in a systematic way. In other words, there are very few investigations which have the evaluations that are being done in a concrete field as their main object of study. Doing metaevaluations can be an original approach for these kinds of studies. A metaevaluation is defined here as the systematic gathering, analysis and assessment of a pre-determined set of evaluation processes. Thus, the (meta)evaluation can be a useful tool for the best understanding and improvement of the evaluation function around a concrete policy sector.

An in-depth review of the concept, aims and functions of metaevaluations is done. It is distinguished from other related concepts, such as evaluation synthesis and meta-analysis. We also identify different types of metaevaluations, define (meta)evaluation criteria for the assessment of evaluation processes, and elaborate on the strenghts and weaknesses of these kinds of studies.

Finally, this paper aims to discuss and reflect on the lessons learned and method of metaevaluation applied in a metaevaluation on gender equality policies in Spain, conducted in 2000-2001.

Meta-evaluation and evaluation synthesis: concept and differences

The concept of meta-evaluation was first developed by Michael Scriven, as early as 1969, although an articulated definition by this same author can be found later on in his “Evaluation Thesaurus”. For him, meta-evaluation is:

“the evaluation of evaluations - indirectly, the evaluation of evaluators - and represents an ethical as well as a scientific obligation when the welfare of others is involved. It can and should be done in the first place by an evaluator on his or her own work; although the credibility of this is poor, the results are considerable gains in validity...[Because] the results of self-evaluation are notoriously
unreliable, however, it is also desirable, wherever cost-justifiable, to use an independent evaluator for the meta-evaluation” (Scriven, 1991, p.228).

The Joint Committee on Standards for Educational Evaluation, in the second edition of the standards, develop meta-evaluation as an accuracy standard to be followed by evaluators. The standard A12 says:

“The evaluation itself should be formatively and summatively evaluated against these and other pertinent standards, so that its conduct is appropriately guided and, on completion, stakeholders can closely examine its strengths and weaknesses” (Joint Committee, 1994: 185).

For Patton (1997: 143), meta-evaluation is the “evaluating the evaluation based on the profession’s standards and principles”.

However, the concept of meta-evaluation may include one of evaluation synthesis, for other authors. For the Evaluation Research Society (Evaluation Research Society Standards Committee, 1982), meta-evaluation was one of the six types of evaluation identified, and it was defined as a way for re-analysing results of one or various evaluations in order to determine what has been learnt about a policy or a program (Chelimsky, 1989). Vedung (Vedung, 1997) defines meta-evaluation as the evaluation of evaluation and feedback meta-processes, although he considers also as meta-evaluation the synthesis of the results from different evaluation studies. In the same manner, Ballart considers both activities to be meta-evaluation: on one side, it “consists on the analysis of the results of various evaluations” and, in the other, it is the review of a type of a broader evaluation research developed in a certain period of time” (Ballart, 1992: 81-82). Although Weiss (Weiss, 1998: 236) points out that when the studies analysed are evaluations, meta-analysis is some times called meta-evaluation, she recognizes that the term (meta-evaluation) should be kept for the first reviewed concept of evaluations of evaluations.

Although the term “meta-evaluation” has been traditionally used for both the evaluation of evaluations and the synthesis of the results of one or various evaluations, we want to clearly distinguish both terms. Interestingly enough, both activities –meta-evaluation as evaluation of evaluations, and evaluation synthesis- are very different in nature.

An evaluation synthesis serves for gathering and summarizing the results of evaluation studies of similar programs or policies. Thus, the focus is on results. Its main aim is not to assess the evaluation processes but to have an empirical base for judging the general findings of the evaluated programs or policies. An special type of synthesis is the so-called meta-analysis which consists on a statistical procedure for comparing findings of different studies, that is, a methodology of research synthesis according to principles of quantitative data analysis. A first definition was introduced by Glass: “meta-analysis refers to the analysis of analyses...the statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the findings” (Glass: 1976 3). In very simple words, an special parameter –effect size-, is estimated for each study analysed, and then the statistical comparison can be made. For Weiss, meta-analysis in the evaluation field could be defined as the “systematic analysis of the results of a set of evaluations of similar programs in order to estimate the overall effect of those programs” (Weiss, 1998: 333). However, because there is no special
procedure for the synthesis or “meta-analysis” of qualitative studies, and due to the fact that most evaluations studies nowadays combine quantitative and qualitative techniques, the use of meta-analysis in evaluation is somehow limited.

This general idea of synthesising and utilising evidence and studies already done has become very popular in the last years, and the evaluation field is not an exception (Chelimsky, 1997; Cook, 1997). However, in countries like Spain, there is still a strong need for empirical evidence and actual evaluation studies, which are relatively new and quite rare. Probably, we need to focus first on getting those studies, and then we could eventually think of synthesizing or meta-analysing them.

In a different manner, a meta-evaluation consists on evaluating evaluation processes. The focus is on those processes—that is, in how evaluations are done—, not on the results or findings of those evaluations. It is plausible for a meta-evaluation to include also the analysis of evaluation results, but just in the sense whether or not they are good findings (that is, if sufficient evidence was gathered, if conclusions were sounded, and interpretations, judgments and recommendations were logically drawn, and so on), and they can be useful or actually utilised for policy improvement, accountability or enlightenment. The interest of the results by themselves, that is in their content, is rather the focus of an evaluation synthesis and not a meta-evaluation.

Thus, while meta-analysis and evaluation synthesis are related to the concepts of evidence-based policies and the accumulation, review and updating of knowledge in a concrete field, meta-evaluation relates to quality control of evaluation processes and to a better understanding of the evaluation function in the policy cycle. And to the accumulation of knowledge in the field of evaluation. Meta-analysis and evaluation synthesis focus on policies. Meta-evaluation does it in the evaluation of those policies.

The meaning of Meta-evaluation and its functions

Meta-evaluation, as defined in this paper—that is the evaluation of evaluation, different from evaluation synthesis—, has a first meaning as evaluation quality control. In a sense, this is a self-reference or “reflexive” meaning, which tries to answer Scriven’s famous question of “Who evaluates the evaluator?”. The concept is thus associated to the control of potential evaluator’s bias and the increase of evaluation credibility. Following this meaning, there have been analogies with the auditing processes, and it has been related to the concepts of accreditation and quality control of evaluations and evaluators (Schwandt and Halpern, 1988). In this role of quality control, meta-evaluations tend to be done individually, that is one evaluation is meta-evaluated at a time and not in comparison with other evaluation processes.

The second meaning or role that a meta-evaluation may play is the one of describing, analysing and assessing evaluation studies or processes. This meaning, which we want to stress in this paper, has been less developed than the first one. Vedung (1997) points out, just as a possibility, that a meta-evaluation might play the role of auditing the evaluation function in governmental agencies. And Rist (1990), Ballart (1993) and Derlien (1998) have done some comparative analysis of the evaluation function in different countries. But we were not able to find many studies analysing or comparing concrete evaluation processes.
In this paper, this second meaning is developed. Thus, meta-evaluation is presented not only as a way for controlling the quality of evaluations, but also as one for studying public policies and interventions, in their concrete evaluation phase. Moreover, we do not want to stress the assessment of evaluation quality, but the assessment of their adequacy and opportunity to the policy and evaluation context, and their contribution to the accomplishment of the evaluation functions. However, we also take into account the first meaning –meta-evaluation as quality control- in order to define the meta-evaluation criteria for that assessment.

**Types of meta-evaluation**

Actually, a meta-evaluation is an evaluation which object of study is a program evaluation, instead of that program by itself. So that, meta-evaluations might be classified and distinguished as any other evaluation.

Thus, in order to distinguish the different types of meta-evaluations, we use in this paper what we have already called the four basic classifications\(^1\): according to the role the evaluation plays (formative or summative); according to the content or the program phase evaluated (design, process or results); according to the moment of the evaluation (ex-ante or ex-postfacto); and according to the agent (internal or external). Applying these classifications to a metaevaluation, instead of a regular evaluation, we have the following types:

**According to the evaluation role**, a **summative** meta-evaluation is done when the role to play is one of recapitulation of the whole evaluation process, that is an overall assessment of the evaluation study or process is done. A **formative** meta-evaluation is used as an instrument for changing the evaluation study during its implementation process; the purpose is one of improving the ongoing evaluation design and implementation. In fact, the Joint Committee standard of “meta-evaluation” is clear about this, stating that an evaluation by itself should be formatively and summatively evaluated; that is, it explicitly recognizes both types of meta-evaluation. Also Scriven, in his *Evaluation Theasurus* points out that a meta-evaluation can be done formatively, summatively, or in both manners; this author stresses the summative meta-evaluation because, according to him, this kind of meta-evaluation offers the client information about the technical competence of the primary evaluator. For him, even if this summative meta-evaluation is not demanded by the client, it should be considered by the primary evaluator as another service to the client, which improves evaluation quality (Scriven, 1991: 230).

Related to this first classification is the one **according to the moment** the meta-evaluation is done. Meta-evaluations can be done **ex-ante** (before the evaluation process is implemented) or **ex-post-facto** (during the evaluation process or when it is already done). Schwandt and Halpern distinguish among provaluation (if the meta-evaluation is implemented before the evaluation is done, that is during the development of the study plan or design), retrovaluation (if it is implemented after the evaluation is done, that is “a retrospective evaluation of an inquiry once it has been carried out”), and monitoring during the study for checking the progress to date and reflecting on “the extent to which

---

\(^1\) They are called basic because every evaluation process can be assigned to one or more categories in all four classifications (Bustelo, 2001). For example, an evaluation might be classified as summative, a process and result evaluation, ex-post-facto, and external.
provaluative considerations have been adequately managed” (Schwandt and Halpern, 1988: 8).

The third classification deals with the agent in charge of the meta-evaluation. According to the agent, there might be internal and external meta-evaluations. Internal meta-evaluations are carried out by the evaluators by themselves: thus, the same people in charge of the evaluation do its meta-evaluation. External meta-evaluations are done by someone not involved in the assessed evaluation process or study. This two types are really different in nature. An internal meta-evaluation means that the evaluation team by itself carries the quality control of the evaluation process (as seen later in this paper, there are some checklists available for this purpose). However, an external meta-evaluation not only implies someone else in the process whose participation gives more credibility to that process, but also means an external recognition and consolidation of the activity of meta-evaluation by itself, because additional resources which go beyond the evaluation are needed. Increasing the credibility of evaluations is one of the main purposes of external meta-evaluations. As a matter of fact, Patton (1997) gives meta-evaluation as an example for combining internal and external evaluation: he proposes the possibility of doing internal evaluations, and at the same time, have someone who externally revise and check the evaluation process. This results in an internal evaluation which has the credibility of external evaluations.

Finally, and according to the content, three types of meta-evaluation -design, process, and results- might be distinguished. It must be clear that meta-evaluation deals with the design, the process and the results of the evaluation and not the program being evaluated. Doing a design meta-evaluation implies the assessment of the evaluation plan and design, and how it was conceived, including its rationale and purposes, objectives, internal coherence, context adequacy, stakeholders identification, strategies for implementation and so on. This type of meta-evaluation should address the analysis of the evaluation context and whether or not the evaluation design is responsive to it; also the cost-effectiveness of the study, and the expectancy of usefulness and utility for the improvement, accountability and enlightenment of the program; if the calendar is realistic and the design includes stakeholder participation, and so on. Process meta-evaluation analyses the evaluation implementation. This type of meta-evaluation includes the analysis of difficulties in the implementation of the evaluation process, calendar and previous plan adjustment, adequacy and effectiveness of the evaluation team work, and so on. Finally, the results meta-evaluation deals both with evaluation out-puts and evaluation outcomes or effects. In this case, the evaluation outputs are the results and the produced information and evidence (findings, interpretations, judgments, and recommendations), which typically are reflected in evaluation reports (although they might be also presented orally). Thus, a meta-evaluation of outputs analyses the quality of those results and information produced and how they are reflected in the evaluation reports, whether the final report is complete and was handed on time and disseminated to the stakeholders previously identified, and so on. However, a meta-evaluation of outcomes deals with the effects an evaluation is supposed to have: learning and better understanding of the evaluated program, implementation of the recommendations given by the evaluation, program improvement, etc. In sum, the effects of an evaluation are directly related to its utility and utilization. In table 1, the four basic evaluation tipologies, applied to different types of meta-evaluations, are exposed:
**Table 1. TYPES OF META-EVALUATION**
Classification according to the four basic evaluation typologies

<table>
<thead>
<tr>
<th>ACCORDING TO THE ROLE OF THE META-EVALUATION</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Formative:</strong> When the meta-evaluation is used as an instrument for improving and changing the ongoing evaluation design and implementation.</td>
</tr>
<tr>
<td><strong>Summative:</strong> When the role to play is one of recapitulation of the whole evaluation process, that is, an overall assessment of the evaluation study or process is done.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ACCORDING TO THE MOMENT OF THE META-EVALUATION</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Ex-ante (provaluation):</strong> Meta-evaluation is carried out before the evaluation process is implemented.</td>
</tr>
<tr>
<td><strong>Ex-postfacto (retrovaluation):</strong> Meta-evaluation is carried out during the evaluation process or when it is already done.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ACCORDING TO THE AGENT WHO META-EVALUATES</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Internal:</strong> They are carried out by the evaluators by themselves, as an internal control (auto-maetaevaluation).</td>
</tr>
<tr>
<td><strong>External:</strong> Done by someone not involved in the assessed evaluation process or study, being an external control.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ACCORDING TO THE CONTENT OR EVALUATION PHASE META-EVALUATED</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Design meta-evaluation:</strong> It focuses on the design, context and structure of an evaluation study.</td>
</tr>
<tr>
<td><strong>Process meta-evaluation:</strong> It focuses the evaluation implementation, that is, how the study is carried out.</td>
</tr>
<tr>
<td><strong>Results meta-evaluation:</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

*Meta-evaluation criteria*

Before elaborating on this point, we want to review what have been said by some authors about meta-evaluation criteria. Vedung just mentioned “quality, readability, fidelity to facts, and other properties” (Vedung, 1997: 41). Scriven suggested the application of his KEC (*Key evaluation checklist*) to the evaluation processes, being this
checklist an instrument for meta-evaluation (Scriven, 1991: 230-231). He insisted on the importance that the meta-evaluator make explicit the values and criteria that he/she is going to use for judging the evaluations. The dimensions in the checklist have to do with the evaluation design, process, findings and results, team, calendar, resources and costs, recommendations and reports, if the context was taken into account and the relevance of the study, if the “consumers” were identified, and if values and criteria were made explicit. Following Scriven checklist, a complete set of meta-evaluation criteria could be set.

Another set of criteria for meta-evaluations are the Standards of the Joint Committee. Actually, these standards are the most common reference for meta-evaluating quality evaluation, due probably to their specificity, clarity, method and structure. As we have already said, in the second edition of the Standards (1994), “meta-evaluation” appeared as a new and specific standard, although in the first edition (1981) the idea of using the list of standards as evaluation quality checklist was already suggested. Thus, each standard could be a meta-evaluation criteria.

Among the first authors who identified some criteria for doing meta-evaluation, were also Schwandt and Halpern (1988), who proposed meta-evaluation as a metaphor of fiscal auditing, and as a way for enhancing evaluation quality. These authors pointed out two main aims for the meta-evaluation: one, to analyse whether the methodology used was adequate (evaluation plans and procedures), and two, whether or not findings were good and right. Following these two aims, we elaborate a little more on each one.

The use of an adequate methodology has to be with the procedures employed for learning about the object of study. In the first place, there should be some coherence among the different epistemological, methodological and technical levels. But, over all, the methodology should be responsive enough to the concrete context and the evaluand, that is, it should relevant to the evaluand and its context. It is crucial that the evaluation criteria are previously established, and the perspectives and rationale that are the bases for value judgments are clearly described. Also, there is a need of technical coherence between the design for the establishment of evaluation criteria, the design for collecting and analysing the information gathered, and the concrete techniques employed for that. We also need to take into account the cost-benefit and cost-effectiveness relationship (the evaluation should be worthy), and the calendar and timing. Techniques should be valid and reliable. Finally, the methodology should allow more or less stakeholder involvement, depending on the approach we are using.

We can also think of what it is to obtain good and right findings in evaluation processes. Essentially, it deals with a logical coherence between findings, interpretations, judgments, and recommendations. Interpretations should be made based on reliable data or evidence systematically gathered. Those interpretations should be the base for judgments. Last, recommendations should be based on judgments, which should be based on interpretations, which should be based on findings. Schwandt and Halpern (1988: 34) mentioned some attributes for good evidence. They are:

Relevance: information, data and evidence should be directly related to the objectives of the evaluation.

Reliability: evidence must be competent and trustworthy.
Sufficiency: there must be enough evidence to draw conclusions within reasonable limits of risks.

Representativeness: data and information gathered must be reasonably representative of the whole evidence that is under examination.

Timeliness: cost and time must be weighted, and sometimes it is not reasonable to gather additional evidence if there is a need to deliver a valid opinion in a specified time frame.

Thus, if the information and data gathered are right (they are relevant, reliable, sufficient, representative, and they are gathered on time) and the analysis is correct and sufficient, the interpretations will be also adequate. Analysis criteria and standards are normally associated with the concrete techniques employed. Judgments should be made according to the evaluation criteria previously established and agreed with stakeholders. Evaluation criteria should also reflect the absolute (what it is consider good and desirable in absolute terms) and relative (what is good or desirable compared to other similar programs) norms or standards which should guide judgments.

Good recommendations should be based on the three previous levels (findings, interpretation, and judgments), and they should take into account the specific evaluand and its context. They must be realist, clear and the evaluation team should allow themselves enough time for establishing them. It is also desirable for recommendations to add an estimation of the costs for their implementation, because this is something that increases the probability of utilization (Chelimsky, 1998).

However, we want to stress here that meta-evaluation might also include other criteria beyond adequate methodology and right findings. Generally speaking, meta-evaluation can also be used for a better understanding of the evaluation processes, and the role they play or might play in the program or public policy evaluated. In a sense, it is way for analysing and assessing the evaluation function –and how it is executed- in a concrete sector, policy, organism or institution. Thus, it is a way for understanding the political processes which enhance, intervene, and arouse with an evaluation process. While meta-evaluation for quality control purposes is better set individually, that is, it is better to meta-evaluate just one evaluation at a time, it is more convinient to analyse various evaluations if the purpose of the meta-evaluation is this one of understanding evaluation processes in the context of public policies.

In table 2, there is a summary of the meta-evaluation criteria already discussed.
### Table 2
META-EVALUATION CRITERIA

<table>
<thead>
<tr>
<th>QUALITY CONTROL purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Using an ADEQUATE METHODOLOGY</strong></td>
</tr>
<tr>
<td>- Coherence between epistemological, methodological, an technical levels</td>
</tr>
<tr>
<td>- Adequacy to the context and the evaluation object –evaluand-. Relevance</td>
</tr>
<tr>
<td>- Previous establishment of evaluation criteria</td>
</tr>
<tr>
<td>- Technical coherence between evaluation criteria, the way in which information is collected and analysed, and the subsequent judgment according to that criteria</td>
</tr>
<tr>
<td>- Cost-benefit and cost effectiveness of the evaluation study</td>
</tr>
<tr>
<td>- Adequacy to the real and available calendar and timing</td>
</tr>
<tr>
<td>- Validity and reliability of the techniques used for collecting and analysing information</td>
</tr>
<tr>
<td>- Stakeholder participation and involvement in the evaluation process</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Obtaining RIGHT AND GOOD FINDINGS</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Coherence between findings, interpretations, judgments and recommendations</td>
</tr>
<tr>
<td>- Good collected findings, data and evidence:</td>
</tr>
<tr>
<td>- √ relevance</td>
</tr>
<tr>
<td>- √ reliability</td>
</tr>
<tr>
<td>- √ sufficiency</td>
</tr>
<tr>
<td>- √ representativeness</td>
</tr>
<tr>
<td>- √ timeliness</td>
</tr>
<tr>
<td>- Good interpretations: correct and sufficient analysis</td>
</tr>
<tr>
<td>- Good judgments: according to and coherent with evaluation criteria, and with absolute and relative norms and standards, they have to be drawn from empirical findings</td>
</tr>
<tr>
<td>- Good recommendations: they should be bases on previous findings, interpretations and judgments, and they should be realistic, adequate and relevant to the evaluand and its context</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>UNDERSTANDING OF EVALUATION PROCESSES purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>(better understanding and knowledge of public policies)</strong></td>
</tr>
<tr>
<td>- Role that evaluation play in public policy</td>
</tr>
<tr>
<td>- Assessment of the evaluation function: Is evaluation integrated in public policies, organisms and institutions?</td>
</tr>
<tr>
<td>- Usefulness of evaluation for the improvement, accountability and enlightenment of public policies</td>
</tr>
</tbody>
</table>

_A Case Study: A meta-evaluation on gender equality policies in Spain_

Eleven evaluation processes were selected for the meta-evaluation study among Spanish regional and national gender equality plans. All evaluations which actually took place between 1995 and 1999 were selected. We decided to select all the existing
evaluation processes, with the condition of being (or pretending to be) an evaluation of the plan as a whole (excluding evaluations of specific programs inside the plans)\(^2\). (Bustelo, 2001).

Following the concept of meta-evaluation that we have discussed in this paper, the evaluation processes by themselves and not their results were the focus of the study. From the meta-evaluation perspective, the purpose of the study was to evaluate those evaluation processes. For that, the following set of evaluation criteria or questions was established:

1. The overall design of the evaluation processes, including their responsiveness to their context, the clarity of their purposes and objectives, the existence of institutional structures and resources for the evaluation, and the utilization of different types of evaluations.

2. Evaluation key elements, such as the stakeholders involved in the evaluation processes, the moments and time when these processes took place, the evaluation criteria established and the methodology and technics used for gathering and analysing information.

3. The evaluation utilization, taking into account the adequacy and usefulness of the produced information and evidence, their communication and dissemination processes, and their utilization and impact both in the gender equality policies and in the women’s agencies (institutions in charge of the evaluated plans).

The logic base of those evaluation questions and for judging the evaluation processes was built around the following six main criteria:

1. The centrality of the evaluation processes in the women’s agency or organism. We think that centrality is necessary for introducing a certain culture of evaluation which leads to the mainstreaming of the evaluation function in public administrations, as a way for improving public policies.

2. Responsiveness to the plan or policy context and clarity of the evaluation purposes. The evaluation take place in an specific and political context and it should respond to that context. The evaluation purposes must be clear and explicit.

3. Clarity and centrality of evaluation criteria (of WHAT is evaluated). The evaluation criteria should determine and guide the evaluation study, being the main orientation for the gathering and analysis of the needed information. That is, the evaluation criteria determine the variables to study and, also, the logic base for systematically judging the information previously gathered and analysed. Moreover, they should be the “organizer” of the evaluation processes: the technics for gathering information should be chosen after the evaluation criteria are defined and not viceversa.

4. Adequate management of evaluation resources, including

\(^2\) Although in some cases, it was not clear they were real evaluation experiences, they were included in the study if the women’s agency considered it as an evaluation of the respective plan.
4.1. a good use of the different types of evaluation,

4.2. the existence of adequate co-ordination structures which allow a reliable and collaborative information gathering

Is extremely important to design and have adequate institutional and co-ordination structures for the evaluation which allow the collection of valid and reliable information. We think this is even more important than the concrete techniques used for the collection and analysis. Having a valid and reliable instrument for gathering information is a necessary but not a sufficient condition for getting the relevant and reliable information for evaluating the plan.

4.3. a good management of time and timetables

There are two issues related with time which are key elements for evaluation processes. First of all, the moment in which the evaluation is planned and started. The evaluation is a rather more powerful tool if it is planned from the beginning and in a continuous manner. Secondly, there is a need for timeliness in the evaluation processes. The evaluation criteria arise in a concrete context, situation and moment: if it takes a lot of time to answer the evaluation questions, they might turn irrelevant or the addressees might not be the same. So that, evaluation responsiveness is clearly linked to the evaluation timetable. The more delays, the less opportunities for being useful.

4.4. enough resources investment in evaluation.

5. Enough elaboration of the gathered information during the evaluation processes. In evaluation, besides gathering and analysing information, that information should be systematically judged in the light of the evaluation criteria previously set. This process should follow in the establishment of recommendations for the improvement of the evaluated plan. Judging and establishing recommendations is what we call “practical elaboration” because, in order to be useful and used, information needs to be elaborated or “translated” into concrete and relevant guides of action. And it is clear that evaluation needs to be useful, practical and action-oriented for the improvement, accountability and enlightenment functions.

6. The existence of good communication and dissemination processes of the evaluation results and reports. This is a necessary condition for evaluation utility and utilization. A first step consists on having a good “product” which reflects well the evaluation process, normally completed and good evaluation reports. In the next step, there are the dissemination processes of those reports, that is, of the evaluation results.

This meta-evaluation allowed us to draw some conclusions about the evaluation function around gender equality public policies in Spain. We think that most of these conclusions could be generalized also to other sectors in the Spanish context. The conclusions were formulated as factors which have influence the adequacy and utility of
the evaluations studied. We briefly expose here, as an example, some of theses conclusions:

1. Lack of clarity in the evaluation purposes: Gender equality policies and plan’s evaluation vs. investigation of the women status

Analysing the evaluation experiences, we concluded that it is necessary to distinguish between doing investigation about the women status and doing evaluation of gender equality plans and policies. Some of the evaluation studied were investigations about the women status, but, clearly, they were not evaluations of the gender equality plans, because there was no direct or formal linkage between the variables studied and the plan being evaluated.

There are many reasons for distinguishing investigation and evaluation (Bustelo, 1999b), but here we want to point out two of them. One related to the object. In this case, the object of investigation is the women status (differentially with men status, that is, the public issue of gender inequality) and its evolution through time. The evaluation’s object, however, is the plans and policies, that is, how the authorities deal with the gender inequality problem and how those public actions influence that public issue of gender inequality. Secondly, there are differences in purposes, being improvement and accountability the main ones for evaluation and knowledge construction for investigation. Obviously, the investigation about the women status contribute to a better policy formulation. But this does not avoid the need for evaluating those policies. Both activities -investigation and evaluation- are crucial for women’s agencies. Moreover, the so-called “equality observatories”, which aim is to collect information periodically about women status, besides their utility for policy formulation, are very useful for evaluating gender equality policies. But the very fact of their existence does not imply evaluation.

2. Lack of a global vision of the public action taken for promoting gender equality: Policy vs. plan evaluation

None of the experiences analised is an evaluation of the global gender equality policy but they evaluate concrete plans approved by their respective governments. The plans intend to promote, co-ordinate and gather all the governmental actions regarding gender equality, so that in a sense, the best way for evaluating the gender equality policies is evaluating gender equality plans. Plan evaluation facilitates a look of the development and results of that plan but does not question the plans’ role or the plans by themselves. Issues related to the plans’ adequacy as the policy main tool, to the actual effectiveness of the gender mainstreaming strategy and its pertinence for the women agencies’ role and functions, are questions that were not addressed in the evaluations studied.

3. Lack of recognition of the political nature of evaluation

One of the reasons for the ambiguity in the evaluation purposes that we have found in the experiences analised, has to deal with this. In some cases the evaluation was considered an exclusively “technical” matter, leaving unattended key elements in the evaluation processes. This circumstance has led, in part, to unuseful or residual evaluation experiences.
4. The perception of evaluation as a secondary function: The important role women’s agencies should play around policy evaluation

In Spain, women’s agencies have been the main leaders and promoters of gender equality policies. Although the key component of those policies is gender mainstreaming (that is, they intend to involve other governmental actors), these agencies keep on being the main actors formulating, promoting, co-ordinating and monitoring those policies. Taking this into account, policy evaluation should be crucial among the agencies functions. However, the evaluation function is still a residual one in most of the Spanish women’s agencies.

5. Need to know exactly WHAT we want to evaluate: the “dictatorship” of the methodology and the techniques

We have found, in the experiences analysed, that it seems that there was not a previous and profound reflection about what to evaluate. Sometimes the WHAT, that is the evaluation criteria, was established and defined through the HOW or the methodology or techniques used, giving priority to this last term.

6. Importance of the institutional and co-ordination structures for evaluation

One of the main problems especially mentioned by the professionals in charge of the evaluation processes, was the difficulty for gathering information from other governmental units which were supposed to participate actively in the implementation of the evaluated plan. The horizontal nature of the gender equality plans means that a good part of the information which needs to be gathered for the evaluation it is not in the respective women’s agency –which is the responsible unit for the evaluation- but in other –quite many- governmental units. This task of collecting information was not an easy one.

7. Importance of timeliness

Most of the experiences analysed have been planned almost at the end of the plan’s implementation period. This has led to serious difficulties in the information collection.

8. A clear deficit of “practical elaboration”

Generally speaking, analysis phases of judging and establishing recommendations were almost inexistent in the experiences analysed. Even, in some cases, there was a lack of a global analysis. If there were a “process” evaluation and a “results” one, there was no overall or final analysis for “putting everything together”. In few cases, there were some final conclusions –although when they existed, they were very short and not global-, but none of the eleven reports contained any recommendation. In sume, it seemed like the evaluation processes were not really finished, losing opportunities for being really useful.

9. Poor communication and dissemination processes

Generally speaking, the communication processes around the experiences studied were quite deficient and evaluation results could have been disseminated much better. Linked to the communication processes it is the accessibility of the evaluation reports.
For the investigation purposes, in some regions we did not have any problem for getting the evaluation reports, but in others we had serious problems, being unable of getting the whole evaluation reports. In general, the accessibility and the public character of evaluation reports is still not clear and there are no criteria for which one ought to be their status. From our point of view, this fact needs to be seriously discussed.

10. A need for a greater resource investment in evaluation

In the meta-evaluation, we could only estimated the evaluation investment through the direct expenses employed. For that, we compared the direct evaluation expenses of the external experiences with two different figures: plan expenses and women’s agency expenses during the plan period. Even, in some of the cases we could not do that, because it did not exist a concrete estimation of the plan general expenses. Logically, when we had those estimates available, plan general expenses were much higher than the women’s agencies expenses, because other governmental units, and not only the respective women’s agency, contribute to the plan. Although limited, after the analysis we were able to conclude that the investment in evaluation was extremely low. Just an example: the most expensive external experience (which was three times more expensive than the following one, in absolute terms), the evaluation of the second andalusian plan, represents the 0.97% of the agency’s expenses (Instituto Andaluz de la Mujer) and only the 0.07% of the plan general expenses. It is obvious that there is a need for a resources increase.

Conclusions

In spite of the difficulties that meta-evaluations might have –especially those related to the accessibility of information and the heterogeneity of the evaluation experiences to include in the study-, we still believe that these kinds of studies can clarify and be useful for the improvement and development of the evaluation function in public administrations. In addition, these studies may also contribute to a higher consciousness about that evaluation function and thus, to an increase of the evaluation culture inside public institutions.

Evaluation practice in Spain, and in some other european countries, is still incipient, although is clearly increasing. We are starting to have some examples and evaluation studies that very different in nature, so that their study, analysis and assessment are needed. In this sense, we have the opportunity for meta-evaluations in other fields, facilitating some improvement of the evaluation function in concrete sectors.

However, although evaluation practice is increasing, we still have a long way to run. Not only we need more evaluations but also we have to do more research about them. We think it is also important to translate evaluation professional practice into scientific terms and to do research which object of study is the evaluation function in the public sector.

Finally, although we have insisted in this paper on the meta-evaluation purpose of description, analysis and assessment of evaluation processes, we also think that evaluations and evaluators should be evaluated. Thus, it is also necessary to develop meta-evaluation with quality control purposes. This arises different and new issues
about who must and can be meta-evaluators, and whether or not accreditation agencies for public administrations are needed.

REFERENCES


