GOBIERNO DE ESPAÑA
MINISTERIO DE POLÍTICA TERRITORIAL Y FUNCIÓN PÚBLICA
SECRETARÍA DE ESTADO DE POLÍTICA TERRITORIAL Y FUNCIÓN PÚBLICA
INSTITUTO PARA LA EVALUACIÓN DE POLÍTICAS PÚBLICAS

GUIDELINES FOR THE DESIGN EVALUATION OF PUBLIC POLICIES

INSTITUTE FOR THE EVALUATION OF PUBLIC POLICIES

2020

GOBIERNO DE ESPAÑA
MINISTERIO DE POLÍTICA TERRITORIAL Y FUNCIÓN PÚBLICA
SECRETARÍA DE ESTADO DE POLÍTICA TERRITORIAL Y FUNCIÓN PÚBLICA
INSTITUTO PARA LA EVALUACIÓN DE POLÍTICAS PÚBLICAS

AUTHORS:

-Celia Casillas

-Laura García

- Mónica Macía

-Javier Rico

Director: Celia Casillas

Gudelines for the design evaluation Of public policies

© Institute for the Evaluation of Public Policies 2020
Secretariat of State for the Territorial Policy and the Civil Service

This Guidelines is the property of the Institute for the Evaluation of Public Policies. It may be reproduced wholly or in part, provided that the source is fully and explicitly acknowledged.

Publisher: Ministry of Territorial Policy and the Civil Service

General catalogue of official publications:
http://publicacionesoficiales.boe.es

NIPO: 277190508

Contents

INTRODUCTION

Public policy evaluation is considered a tool for improvement and learning of public policies and for accountability in government action. Within the different approaches to evaluation, comprehensive assessment[1] considers evaluation to be a process that combines the political-strategic analysis of public action with the analysis of its operational aspects. It also takes into consideration the entire life cycle of public policy. This is because the results of public policies cannot be separated from the consequences that deficiencies in their design or implementation may have on said policies.

The Institute for the Evaluation of Public Policies (hereinafter, IEPP), regulated by Royal Decree 307/2020 of 11 February, which establishes the basic structure of the Ministry of Territorial Policy and the Civil Service, is the benchmark organisation of the General State Administration Services (AGE in Spanish) with regard to public policy evaluation. It promotes the culture of evaluation of public policies, the formulation and dissemination of methodologies, and the training of public employees on this subject, in coordination with the autonomous body INAP.

The practice of evaluation is a systematic process that requires knowledge of multidisciplinary techniques and tools, as well as a methodology that gives internal validity to the evaluation. The focus of comprehensive assessment requires us to follow an evaluation methodology that encompasses the entire life cycle of public action: design, implementation, and results and impacts. The National Agency for Evaluation and Quality (AEVAL in Spanish), the predecessor to the IEPP, has applied this comprehensive approach to its evaluations and has developed its own methodology which is described in several documents that have been widely distributed, especially its practical guidelines for evaluation design and execution with the AEVAL approach, 2015.

The Institute for the Evaluation of Public Policies seeks to ensure a high-quality evaluation practice that is based on defined processes and recognised methodologies, and to provide tools that assist in the evaluation of any policy, plan, or programme. These include specific guideliness on the different dimensions of a comprehensive assessment, aimed both at evaluators and the managers or public officials in charge of commissioning said evaluations[2] .

---

[1] For a general overview on this approach to public policy evaluation, see the Guía práctica para el diseño y la realización de evaluaciones de políticas públicas Enfoque AEVAL [Practical Guidelines for the Design and Execution of Public Policy Evaluation AEVAL Approach]. (AEVAL, 2015).

[2] The Institute for the Evaluation of Public Policies has also published the following guideliness: Guidelines for the Evaluability Assessment of Public Interventions (Institute for the Evaluation of Public Policies, 2020), Guidelines for Evaluating Public Policy Implementation (Institute for the Evaluation of Public Policies, 2020) and the Guidelines for Evaluating Public Policy Results (2020). All the guideliness are published on the website of the Ministry of Territorial Policy and the Civil Service in the Institute section: https://www.mptfp.gob.es/portal/funcionpublica/evaluacion-politicas-publicas/Guiasevaluacion.html

The final goal of any public programme or policy is to achieve a series of satisfactory results and the critical element in obtaining them is usually a design that is tailored to the problem and requirements. Experience has shown us that defective design and interaction with the relevant context is one of the most common causes for a lack of results in public interventions. Design evaluation allows us to identify whether a public intervention is correctly designed, if it approaches the problem behind the public action correctly, if there is a causal logic or a consistent cause-effect relationship, and if said design is in line with both the problem and the assessment, as well as with the underlying internal logic between strategic objectives, operatives and activities. This is independent of the time and development of the evaluated intervention, as the design evaluation has uses, however different, that range from the moment of planning itself to the end of the life cycle.

The goal of this Guidelines is to offer some simple guidelineslines for assessing the design of public interventions with two different target groups. On one hand, we have the persons or bodies in charge and the managers of the interventions who may thus have a general overview of the dimensions and contents of the design evaluation. On the other, we have the evaluators, following the design evaluation process, with examples and techniques.

The first part of this document, “General Questions” provides a simple response to basic questions on design evaluation in a question and answer format: what it is, what does it consist of, why is it recommended, and how to perform a design evaluation.

The second part, “Methodology of Design Evaluation” delves into the details of the analyses required to perform the evaluation: the formulation of the intervention, the design of the intervention, as well as the design evaluation criteria and the evaluation matrix of questions and assessment criteria.

PART ONE GENERAL QUESTIONS

What is design evaluation?

The evaluation of policies has focused essentially on their implementation and on the results-effects of public policies-programmes (a positivist outlook). This is due to the position occupied by evaluation in the classic life cycle of public policies as defined by Lasswell and other authors, which is none other than the last stage once the intervention has been implemented. Additionally, it was deemed that the analysis of the design of interventions fell under the planning and definition of public policies.

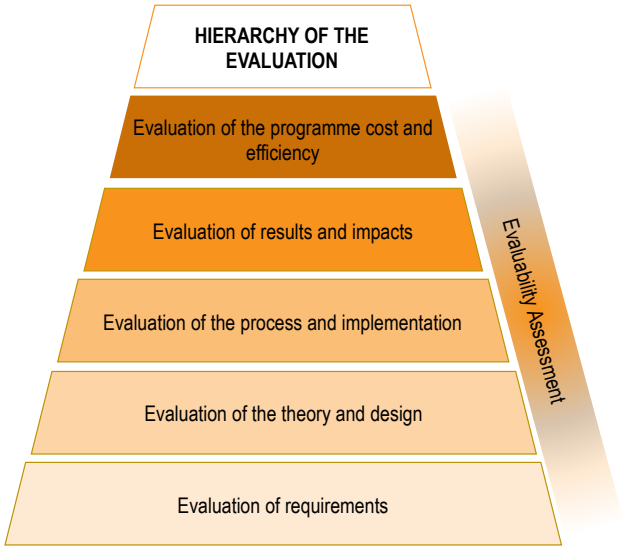
Gradually, the field of evaluation has started to pay greater attention to the design evaluation of public policies or programmes. As a matter of fact the theoretical-practical corpus of evaluation continues to refer to the relevance of design in achieving positive results in public action. Nevertheless, and in spite of the importance of the accurate design of any public intervention (both for the success of the intervention as well as for the evaluation itself), it is only recently that it is being considered as an independent area that deserves conceptualisation and in-depth reflection.

Therefore, until now design evaluation had focused traditionally and primarily on identifying or demonstrating conceptual or pre-existing design problems when evaluating an implemented intervention or its results, by means of a logical-formal analysis of the programme with regard to the problems it sought to resolve (AEVAL, 2010).

The most relevant references to public policy design as a standalone analytical space were made by Rossi, Lipsey and Freeman. For these authors, design analysis focuses firstly on the problem (detected requirements) that the policy or programme attends or seeks to attend to; secondly, on the process of formulation and design of the intervention; and thirdly, on its intrinsic rationality, that is to say, to what extent the design of the intervention serves its goals (Rossi, Lipsey and Freeman, 2003).

The importance of design evaluation for the overall evaluation is brought to light when the hierarchy of evaluation and its relational nature or the relations between the different types or elements of evaluation are established:

Figure 1. Design evaluation within the hierarchy of evaluation. Source: Author's own based on Rossi, Lipsey and Freeman (2003).



Thus "(...) *we can think of these evaluation building blocks in the form of a hierarchy in which each rests on those below it. The foundational level of the evaluation hierarchy relates to the need for the programme. Assessment of the nature of the social problem and the need for intervention produces the diagnostic information that supports effective programme design, that is, a programme theory for how to address the social conditions the programme is intended to improve*" (Rossi, Lipsey and Freeman, 2003).

From this perspective the importance of design evaluation lies in the fact that there is little sense in focusing exclusively on the implementation or the results of the intervention, given that both evaluation types or questions are related to the former or are located at the upper level. The most important of all of them is the analysis of existing needs and the design of the intervention.

It is worth pointing out that for these authors, design evaluation is distinct from the evaluation of the requirements but there is a link between the two, given that design gives concrete shape to the resolution of the pre-existing problem, and the rational nature of the design. A concept that is not passive, as we shall see later.

For Bueno and Osuna (2013) design evaluation is "*that which analyses the rationality and coherence of the intervention; compares the veracity of the assessment which justifies it; judges the definition of the goals proposed with regard to certain parameters; examines the correspondence between the reality of the context in which it seeks to intervene and the (explicit and implicit) goals that the political strategy seeks to achieve*".

For the purposes of this Guidelines, design evaluation is defined as

> ***the activity that performs a systematic analysis of the problem that justifies a public intervention (its characteristics, its assessment on the basis of underlying requirements, its identification as a public problem), the choice of the intervention to solve the problem within the existing social, economic and institutional contexts, as well as the conceptualisation and design of said intervention, based on causal logic (theory of change).***

Design evaluation within comprehensive assessment

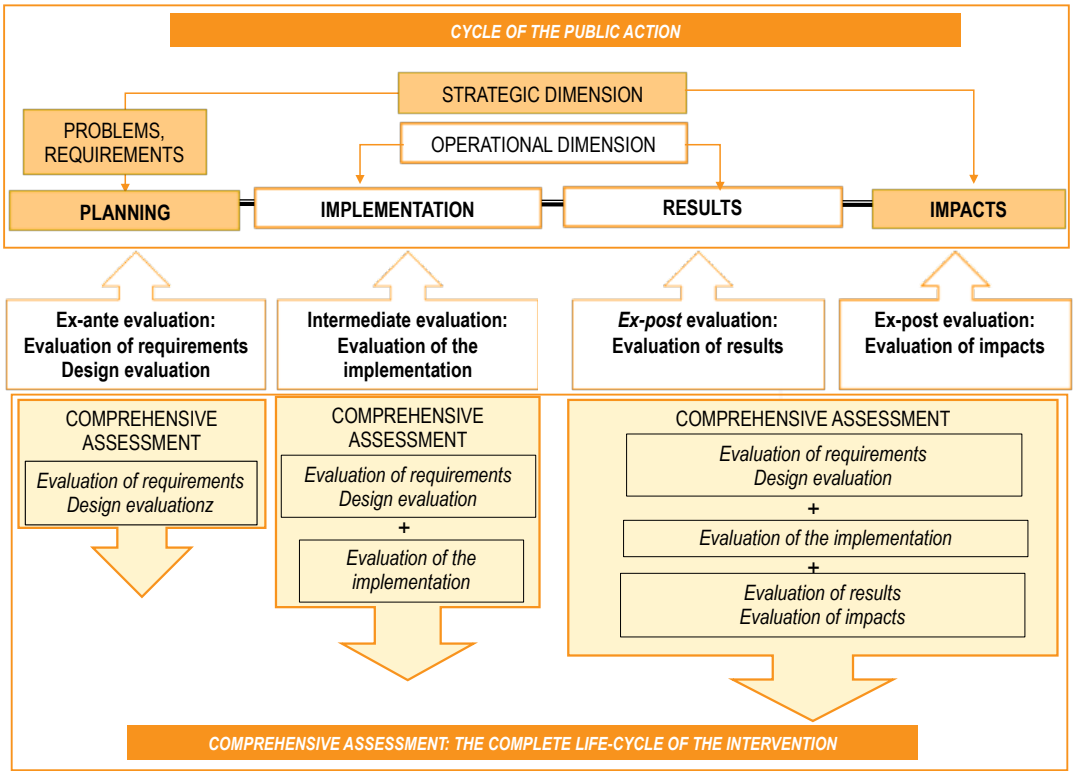
Comprehensive assessment considers that public policies are action processes characterised by their complexity and inter-connections with public problems, which require an approach that fully includes the problems as well as their solutions as opposed to the fragmented and decontextualised analyses offered by classical evaluation.

Given that the premises or assumptions of classical evaluation are not always present, comprehensive assessment takes an approach that encompasses several problems that incorporate *"elements of complexity during the different stages of the process" (Grau-Soles et. al, 2011)* and which considers all related politics. It emphasises the political aspects of public action as a means to comprehend policies (multi-sector, multi-level, which address complex problems where multiple stakeholders with opposing interests interact) (AEVAL, 2015).

Thus, this focus integrates the political-strategic analysis of policies -which includes problems, assessments, stakeholders, intervention theories and hypotheses, formulation and formalisation of the intervention and its impacts- with the analysis of the operational aspects of its rollout, that is to say, its objectives, resources, processes and intermediate and final results.

Evaluation design is also a definitive feature in comprehensive assessment and it addresses the contingency and adequacy of the organisations whose features may affect the achievement of the planned objectives and finally, the inability of the government as sole stakeholder to solve public problems that are not foreseen in the standard evaluation (Ruiz, 2015).

Figure 2. Cycle of public action and comprehensive assessment. Source: Author's own based on AEVAL (2015).



What is design evaluation?

Design evaluation has several uses, depending on the stage within the life cycle of the intervention when it is performed, as it may also dictate the extension, depth, and possible evidence with regard to design, on the basis of the aforementioned comprehensive approach of the evaluation.

If it is undertaken on a fully implemented programme or policy (***ex-post evaluation***), design evaluation will allow us to display the observed results and impacts as well as to identify the factors that play a role in achieving said results, to detect the possible gaps in the underlying theory of change and the interaction of the intervention within a given context, questions that are plainly relevant to design evaluation. The analysis will logically be more in-depth, given that it is possible to obtain evidence of the relation between the problem, the design, the outputs, and the results and impacts, as may be observed in the figure below. It will be possible to judge if certain mediocre or deficient results are due to a design error, or implementation deficits or problems, a problem definition that is lacking, the response of the target population to the programme, or to any other factor.

Figure 3. Chain of causal relationships. Source: Author's own.

Design evaluation at this stage of the cycle is useful within a feedback scheme, for a redesigning or even survival of the programme, although this circumstance is conditioned by the role or importance of evaluation within the public policy formation process.

In the case of interventions at the ***implementation stage***, design evaluation may be useful for an early redesign of the intervention as it may lead to a more efficient early detection of design errors. It is especially useful in the case of interventions for dynamic problems.

Finally, an ***ex-ante*** design evaluation when the intervention has not yet been implemented can make it possible to undertake a full redesign and make significant savings in terms of all possible costs. At this stage of the cycle, design evaluation allows us to identify the causal logic or the underlying theory of change, as well as the problem and the identification of the potential target population, and how they are affected by the problem or requirement.

What does design evaluation consist of?

As mentioned earlier, design evaluation is the analysis of the different dimensions of the planning of a public intervention which allows us to draw conclusions, based on evidence, with regard to its different components.

1. Formulation.

All public programmes or policies must seek to resolve or mitigate a problem. It consists of different elements that influence decision-making on how to approach this problem or requirement taking into account the context and actors.

- The ***problem*** that justifies a public intervention. The analysis of the problem constitutes one of the most critical elements in design evaluation. It involves ascertaining the problem, the causes behind it, its characteristics and etiology, the effects it has, who is affected and how. The evaluation must also analyse how the problem has been diagnosed and existing needs, and if the assessment is adequate and consistent. This also involves correctly identifying and profiling the target or potential target population, as well as establishing different degrees of requirements.

Additionally, design evaluation must also determine how the problem under consideration enters the ***public agenda*** and becomes a public problem that requires an intervention. Occasionally, the publicly defined problem is influenced by interest groups or lobbies, government interests, stakeholders, or managers, all of them with opposing interests most of the time.

On most occasions, interventions act within a sphere where one or more public policies or programmes are already producing a certain effect. The analysis of how the intervention operates within a certain context and the effects produced are elements to be considered when analysing the problem and the requirements within the framework of design evaluation.

- The ***choice of public intervention*** to cope with the problem constitutes the second element of this definition of design evaluation. Faced with certain given requirements, there may be a range of intervention possibilities, based on their approaches and their manner of solving the problem. It does not just involve analysing the designed intervention but also assessing it with regard to considered alternatives.

- The current ***context***. The social, economic, institutional, regulatory, and administrative context in which the interventions take place is decisive, not only with regard to the correct design of the interventions, but also with regard to their potential effects. Experience in evaluation has repeatedly demonstrated that interventions whose complexity does not consider existing contextual factors may be theoretically well-conceived but are generally doomed to failure. In the same regard, evaluations usually demonstrate that successful interventions with regard to specific contexts, moments or countries need not necessarily function in others, an issue related to the repeatability of interventions or their external validity.

2. Design of the intervention.

The design that gives shape to the chosen alternative, constitutes another central component of design evaluation. There are two design elements in any intervention, a more operational one called theory of action or implementation, and a more strategic-reflexive one called the theory of change.

- The ***theory of action*** or implementation reflects the hierarchical unrolling of the intervention into general or strategic objectives, specific objectives and operational objectives, as well as the definition of the activities and measures to be taken for achieving them. The evaluator's analysis of the theory of action involves determining the strategic and operational objectives that the intervention seeks to achieve and their internal consistency with the context. It also involves analysing the means-goals relationship between the objectives and the mechanisms or actions to achieve them, which implies judging the adequacy of the latter to the former. The theory of the process, which takes an in-depth look at the workings of the intervention at two levels (plan regarding the use of the service and the organisational plan)from a theoretical point of view can provide evidence in this regard (Ivàlua, 2009).

- The ***theory of change*** or logical-causal relationship is a critical element of the design evaluation which allows us to explain how and why the intervention will be able to solve the problem or produce certain effects. The analysis of the theory of change attempts to evaluate the underlying bases of the internal cause-effect logic of public intervention, its presence in the intervention, consistency, and reliability with regard to experience and current knowledge. It allows us to judge the consistency and quality of the theory of the intervention, both with regard to how the requirements or problems of the target population are faced, and the contribution of the specific measures or activities through the allocation of resources or inputs and the outputs or services offered. Additionally, in the case of an ex-post evaluation, whether they have been deviations in the intervention as a result of the design mechanisms themselves, the process of implementation, the changes to the context, the effect of incentives and disincentives, etc.

3. Criteria and matrix of the design evaluation.

This dimension includes the methodological elements of the evaluation in performing the design evaluation.

Evaluation criteria. The criteria provide benchmarks (yardsticks, standards, principles, etc.) to obtain useful information in order to assess the evaluation. The criteria of suitability, relevance, internal coherence, external coherence and complementarity are especially important in design evaluation.

The figure below summarises those that are frequently used in design evaluation.

Figure 4. Evaluation criteria in design evaluation. Source: Author's own.

Criteria associated with design evaluation	
Suitability	The degree to which the set of measures that constitute the intervention is directed towards solving the existing problem or requirement within the context in which the problem is generated.
Relevance	Importance of the measures within the wider framework of the policy. Correct sizing of the measures.
Coverage	The section of the population that benefits from the policy or programme, in comparison to all the people affected by the problem.
Internal coherence	Consistency between the objectives of the intervention and the set of measures that have been designed and implemented.
External coherence	Actions performed in the area of other policies or dimensions with the aim of solving certain aspects unforeseen by the intervention.
Complementarity	The degree of alignment between two or more policies or programmes that address a problem.

The ***questions*** are the basic unit of research, which may be defined as the queries and hypothesis to be confirmed which allow us to execute the evaluation. The list of questions and their associated criteria, as well as the sources of information, measurement indicators, techniques and tools that are included in the ***evaluation matrix***, which constitutes the tool that logically unites all these elements. The evaluation criteria and questions are included in the matrix and contain the focus and scope of the evaluation.

This triangular approach is a requirement of comprehensive assessment, as it considers all the theoretical-scientific perspectives that are considered relevant and useful for evaluation. For this, all types of ***techniques and tools*** are used. This Guidelines mentions some of the most commonly used ones.

How is a design evaluation performed?

Design evaluation is a process of evaluation which focuses especially on the conceptualisation of public action, its internal logic, and elements of its implementation in order to solve or improve a need or problem.

The process unfolds on the initiative of the manager or person or body in charge of the intervention, whose decision is usually reflected in an initial document that contains the analysis of the commission. The process is carried out by analysing the listed dimensions and ends, as we have mentioned earlier, with the evaluation report which contains the findings, conclusions, and recommendations.

The duration of the evaluation will depend on the complexity of the intervention and its nature, characteristics, and conditions, which include the resources allocated for the evaluation.
The process concludes with an evaluation report, which must describe the result of the investigation, the different analyses performed and the findings obtained, usually following a structure based on the evaluation queries used and their associated criteria. The final report must include a conclusions and recommendations section, always based on the obtained evidence. If we think of evaluation as another public intervention, then we may close the cycle with a follow-up of the evaluation[3].

[3] The recommendations of the aforementioned AEVAL Guidelines 2015, may be followed when drawing up the report.

PART TWO. METHODOLODY OF DESIGN EVALUATION

Traditionally, the life cycle of a public policy is deemed to have five general stages: identifying and defining the problem, formulating alternatives, adopting the best alternative, implementing the alternative, and evaluation. In a correct planning, this entire process must be apparent.

Within the process of the comprehensive assessment, one of the stages consists of the analysis of reconstruction of the intervention logic[4]. The goal of this stage is to acquire in-depth knowledge on the target of the evaluation, unravelling the internal line of argument of the evaluated intervention on the basis of its different components:

- A component of the ***formulation of public interventions***. This element is a part of the identification of the problem to be solved and the selection of alternative solutions, as well as the context of the intervention.

- A strategic causal component, which is the ***theory of change***, the hypothesis or the causal logic of the intervention. The theory of change refers to how the intervention seeks to generate the required changes and different stages to achieve the intermediate results and the expected final transformation. Later we shall analyse the theory of change and its different notions.

- An operational component, the ***theory of action*** or implementation. It refers to the sequential structure of the plan or programme with regard to goals, activities and resources that lead to the outputs that will generate the results and impacts that are awaited from a public intervention.

  "The goal (...) is to have as much knowledge as possible of the problem behind the intervention, its causes and effects, the alternatives chosen to solve it and the instruments that are required (...) and finally, the design of the intervention itself" (AEVAL, 2015).

This section looks at the analysis of each element in design evaluation, the criteria and the queries to be answered in the analysis, and the main tools to be used.

It is important to remember that when undertaking the design evaluation of a public policy consisting of different plans, programmes and diverse measures, which have not been noted in a single document, the instruments for evaluating said public policy and the methodology detailed in the following paragraphs must be built and applied to each policy section. The analysis used by the evaluator to make an assessment will be based on the sum of these different plans, programmes and measures that are included in the public policy, taking into consideration the relevance and unique scope of each of them.

[4] This guidelines takes an AEVAL approach to monitoring the execution of an intervention which arranges the evaluation process in a logical and rational manner into the following stages: Analysis of the Commission, Intervention Logic, Evaluation Design, Fieldwork and Analysis, Drafting the Evaluation Report, Communication and Dissemination of the Evaluation and finally, the Monitoring of the Evaluation. The description of the stages may be consulted in the aforementioned AEVAL Guidelines 2015.

Example of public policy evaluation: Administrative burdens in company creation (AEVAL, 2012).

The target of this evaluation is the set of measures launched by the General State Administration since 2007 to reduce the paperwork, costs and time required to establish a company, from the point of initiating the paperwork to obtaining the license to perform the activity. Therefore it deals with a multiple target, and not a single programme.

– ACTION PLAN FOR REDUCING ADMINISTRATIVE CHARGES (PARCA in Spanish)

· Digital management...

"PAQUETES CONSEJO MINISTROS"	Fecha de aprobación	Número de medidas
1er Paquete	28 de junio de 2008	11
2º Paquete	15 de agosto de 2008	70
3er Paquete	18 de abril de 2009	78
4º Paquete	24 de diciembre de 2010	88
5º Paquete	14 de mayo de 2011	14
6º Paquete	17 de septiembre de 2011	22
TOTAL		283

– Aspects of LAW 2/2011. ON SUSTAINABLE ECONOMY:

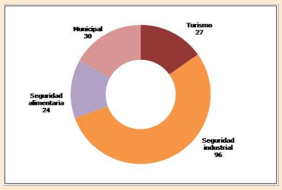
· Implementation of affirmative silence

· Modification of the Local Government Law 7/1985

– TRANSPOSITION OF THE SERVICES DIRECTIVE:

· Effects on different regulations at different government levels and scope

· Signed declaration



– DIVERSIFICATION IN THE REGULATIONS ON THE CREATION OF COMPANIES:

· RD-L 13/2010
· Law 7/2003 SLNE

ANALYSIS OF THE FORMULATION OF THE INTERVENTION

The design evaluation of any public intervention must commence with an analysis of its formulation, consisting of the analysis of the public problem, its entry into the public agenda and how it is defined or configured, the choice of the alternative adopted to solve it; and the context in which the formulation is produced. These analyses allow the evaluator to respond to queries on the suitability, relevance, coherence, and complementarity of the intervention and to make an evidence-based assessment.

Analysis of the public problem

The definition and analysis of the problem that justifies a public intervention is probably the most crucial element of the public policy cycle. The terms under which a problem is considered, its characteristics, the existence of requirements and its causes and effects are basic aspects and elements present prior to the design stage of any public programme or policy, to the definition of its strategic and operational goals and the set of activities, processes, resources and actions involved. The genesis of all interventions. All solutions are "*part of the search for the problem*" (Wildavsky, 1979), therefore, all interventions (solutions) must be based on a correct identification of the problem. An incorrect definition of the problem may lead to more errors than if a problem is well-defined but the solutions are incorrect (Dunn, 1981).

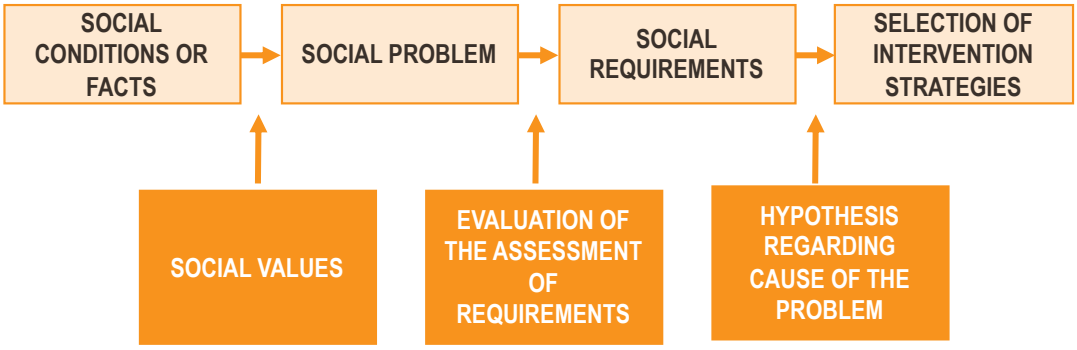
Defining the problem

Public problems are social and political constructs. Problems are often deemed to be objective, when in reality, they are not. There is also no guarantee that a situation that affects certain groups will become a public problem. In order to qualify as the latter, it must enter the public agenda through a process where the social nature of the problem must first be recognised, by leading to certain objective deficiencies in society, and secondly, when stakeholders that wield the power classify said deficiencies as a public problem and it becomes institutionalised. The evaluator must therefore pay special attention to the terms under which a problem is defined and enters the public and government agenda.

As mentioned earlier, some authors draw a distinction between the evaluation of requirements and the evaluation of design (Rossi, Lipsey and Freeman, 2003).

It is possible to establish a sequential logic in the analysis of the problem, similar to Alvira (1991) where social conditions and the institutional, economic, and social context create a certain reality that hinders the full development of citizens, businesses, collectives, or institutions and generates certain social requirements. It is another matter whether said requirements achieve the status of a social problem that is included in the public agenda and which requires public action, and the terms of said public action.

Figure 5. Process of analysis of the problem, assessment of requirements and selection of the intervention strategy. Source: Author's own based on AEVAL (2015).



In spite of its importance, many policies or programmes do not provide a correct assessment or description of the problem, or it is incomplete or partial, or there is no clear identification of the affected population, the intensity or variability of its effects or the requirements based on the group of factors that give rise to them.

A social problem may be defined as the set of *"existing deficiencies in a certain population group, which constitutes a gap between what is desired by society and its reality (...) a population may have an infinite number of possible problems to be resolved"* (Lima Facio and Aguilar Astorga, 2011). But the issue faced by the designer of public interventions as well as the evaluator is, as we have mentioned before, is that the problems are political and social constructs that are dynamic and changing, and where defining the problem goes beyond a simple conceptual definition, as on the basis of this definition, the intervention may have excessively limited or restricted goals, thus underrating important effects or losing the opportunity to select more suitable policies (Moore, 1993).

Among the elements that must be correctly identified in the design evaluation when assessing the problem are:

- ***Its etiology or cause(s).*** An assessment must explore the real underlying causes behind the status of the problem or requirement. It must also make a detailed study of how these direct causes are linked to the current context and other possible factors that, without being the direct or unequivocal cause, modify, characterise, or influence said problem. Problems often have multiple causes.

  It is rare that the contexts in which social problems and requirements exist, there is no public intervention, programme, or public action. These actions condition the problem as well as its effects, with regard to mitigating, exacerbating, or characterising it. The analyses must therefore take into account how public actions influence the problem (external coherence or complementarity of public policies with regard to the evaluation criteria). Additionally, design errors or errors in implementing developed interventions should not be considered as the "problem" in and of themselves. Therefore, inattention, lack of coverage in terms of public assistance or specific public service do not constitute the problem.

Another basic premise in the analysis of the problem in the design evaluation stage is to ***define the target population*** that suffers from the problem or has the social requirement. It must be clearly established who are affected as well as the affected sectors or collectives (affected population). The term target population may be somewhat misleading as it does not exclusively mean the affected population or one that directly suffers the consequences of the social problem or requirement, but all causal elements that lead to the problem. Thus for example, in juvenile delinquency, the target population is not composed solely of juvenile delinquents but also all the realities that induce delinquency. In the case of poverty, the target population does not constitute only of poor individuals, but also all the socio-economic institutions that are responsible for the phenomenon of poverty.

Mistakes to be avoided

Underdefining the problem or sub-optimisation*. (Bardach, 1993). The analysis of the problems should not just consist of the lesser question or problem but also all that give rise to it. That is to say, evaluation must not limit itself to an aspect or problem if it is derived from a bigger problem. Even when the intervention, for various reasons, is limited to only a part of the problem, its analysis must not ignore the general question or problem.*

*The **predominance of the perceptions** or focus of the evaluator or designer with regard to the problem, or ethnocentrism. There is no single focus and the predominance of one over the others should be avoided. This risk usually involves reducing the level of importance of certain problems or aspects that may be much more relevant for other collectives or even individuals. Conversely, it may also involve exaggerating or making certain problems seem over-important (Bardach, 1993).*

The ***intensity or magnitude of the problem*** or the social requirements that it provokes. Public problems never have the same magnitude, rather they affect different collectives or target populations with dissimilar intensities. Inequality, for example, is a problem that displays different intensities which in turn are due to different causes such as extreme poverty, lack of economic and educational resources, the workings of the labour market, etc.

The ***possible definitions of the problem*** or its nature. There are no unequivocally defined problems (Wildavsky, 1979). Both the complexity of the environments in which they occur and the definition of the problem itself condition the creation of the prospective instruments and objectives of the public decision (Aguilar, 1993). The groups and stakeholders that have the capacity to influence the definition and configuration of the problem do not only exert their influence on the decision, but also condition the solutions chosen.

The ***stakeholders*** involved or who must be considered with regard to the public problem, their interests, and strategies: the businesses, public institutions, collectives or groups of interest or lobbies. As mentioned earlier, the stakeholders or groups of interest are able to define the problem according to their own perceptions and interests.

- The ***dynamic nature of the problems***. Complex societies are in a state of constant evolution and new public problems continuously arise while others are mitigated. Additionally, the public interventions themselves, in this area of intervention or in others, tend to provoke changes to the situation.

- The ***consequences or effects*** that cause the problem. Determining, as far as possible, the consequences of the problem envisaged in the intervention. Normally, they must be logically related to the causes.

- ***The duration of the problem***. Whether the duration is continuous or discontinuous, it has effects on the relevance of the problem: highlighting older problems or a lack of knowledge or information in the case of new problems, etc.

Ultimately, it deals with finding a definition of the problem that is increasingly better-expressed but also demarcated.

An approach that is of interest when analysing the problem is one that focuses on the question: *What are the problems for the citizens?* This focus influences the policy design to a certain degree and highlights the identification of the problem rather than its resolution. From this perspective, policy design -and the assessing approach to the problem- is characterised by an attempt to structure the problem, where a problematic situation becomes a clear and correctly expressed definition of the problem which leads to a solution.

The key question in any public policy design and therefore its evaluation, is how to evolve in a responsible fashion from an intractable, unstructured or less-structured problem to one that is better arranged or controlled. This transformation takes place in four processes or stages, that may be expressed as questions (Hoppe, 2018): problem detection ("*Why is the situation a disaster or a problem?*"); problem categorisation ("*What is the gap between the problematic situation and a more desirable situation?*"); problem breakdown ("*Is there a potentially salvageable gap?*"); and problem selection ("*Where exactly are the opportunities for improvement or to mitigate suffering?*"). Although this approach belongs to the design or planning of public interventions, it may also be used in design evaluation, as it also allows us to structure the analysis and compare in order to see if the problem characterisation or definition is adequate.

AEVAL evaluations perform an analysis and breakdown of the problem under the terms listed in the section dedicated to the analysis of the intervention logic. Thus for example, the document "Evaluación de las medidas de racionalización y mejora de la gestión de la incapacidad temporal" [Evaluation of the measures for streamlining and improving the management of temporary disability] (AEVAL, 2009) performs a full characterisation of the problem and analyses the framework of the phenomenon in comparison to other European nations, delving into its causes and collectives, branches of activity, individual characteristics of workers who are furloughed for common contingencies and the factors that condition the occurrence, re-occurrence and duration, so we may precisely ascertain the problem and its scope.

Statistical Techniques

In order to perform this analysis, all available indicators and different statistical techniques are used. For example, statistical techniques such as CHAID and AID analyses, linear and logistic regressions. CHAID and AID are multivariate statistical techniques where it is possible to isolate a dataset according to a series of independent variables and thus examine the behaviour of another variable to be explained. Regression analysis are of great importance for learning about and identifying the variables, factors or elements that explain a fact or reality. They are described further in the section on tools and techniques.

Another example of problem identification and characterisation is the document “Evaluación del programa de ayudas a las actuaciones de reindustrialización (REINDUS)” [Evaluation of the programme to aid reindustrialisation actions] (AEVAL, 2011), which analyses the structural characteristics of the Spanish industrial sector, its link to regional imbalance and the specific problem of industry.

The entry of the problem into the public agenda

The configuration of the public agenda is how problems and alternative solutions gain or lose the attention of the public or the elites (Brikland, 1997). Even when a question gains public attention, it does not necessarily become a public problem, as discussed before. Groups and stakeholders fight or compete to ensure that their description of the problem is predominant and to have their approach to the problem accepted. Its entry into the public agenda is an essential aspect of design evaluation, along with the analysis and characterisation of the problem in the most objective terms possible.
Establishing methodological guidelineslines to analyse the public agenda in design evaluation is a complex affair, given that the specific object of evaluation conditions the approach to it. How an evaluator may approach this analysis is clearly influenced by the nature of the questions formulated. In any case, there are certain guidelineslines for analysing the entry into the public agenda and definition as a public problem:

- On one hand, to analyse the role and capacity of the different stakeholders and especially the ***interest groups and lobbies***, to include their perspective and proposals in the definition of the problem. By using different qualitative instruments or tools, the evaluator must examine the role of the stakeholders in defining the problem. Here, a useful tool is the documentary research of the previous record and position of the different groups with regard to the problem, open or semi-structured interviews with stakeholders and lobby groups, apart from the designers and managers of the public actions; the design and planning documents of the programme or policy; surveys of groups and target populations; discussion groups or NGT , which shed light on the different underlying discourses with regard to the problem, the maps of stakeholders or their contributions to public information processes.

Mistakes to be avoided

*The **predominance of public opinion** in defining and characterising the problem (Bardach, 1993).*

*Prevent the **problem from defining the public policy solution** (P.J. May, 1993). It leads to an incorrect classification of the problems, to applying superficial solutions (policies or programmes that initially appear to be solid but do not resolve the issue) and recommends initiatives with top-down hierarchies.*

***There is a solution to every problem**. Not every problem has a solution nor is there always a better alternative therefore, they (solutions and alternatives) must all be suitably considered.*

- The role of ***government executive bodies and public bureaucracies*** in defining the public problem. Defining the problem does not always consist of an approach or analysis by public bureaucracies, although they play an important role. As a matter of fact, while numerous stakeholders are involved in the consideration of a situation as a social problem, there are fewer participants when defining it as a public problem, generally consisting of government experts, interest groups and public officials. Political-bureaucratic interests must be analysed in the design evaluation. At this level, the concerns mentioned in surveys, the observation of citizens' preferences and their behaviour represented statistically may condition the analysis of the problem or the design of the interventions. There is also the tendency to trust certain types of analysis, such as political-economic analysis, cost-benefit and cost-effectiveness analyses, theories on public finances and an entire set of approaches or tools that, while adequate, do not account for all possibilities of analysis.

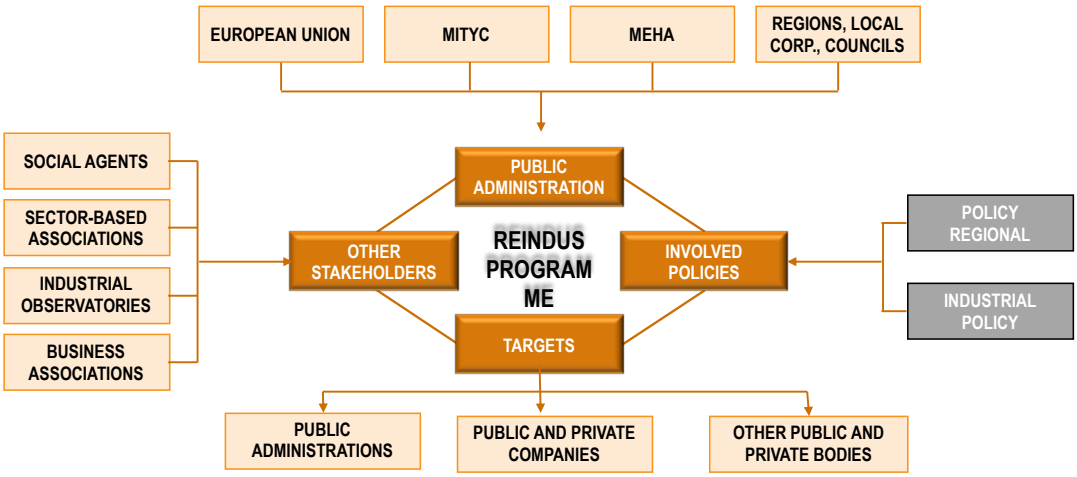
- ***Political interests and the political context***. Public programme design may be conditioned by the programmes of political parties, electoral cycles and considerations with regard to winning, or retaining popular support or political power.

The social realities of governmental action are quite complex and citizens, stakeholders and groups of interest have different beliefs, approaches, needs and interests, whose structure must be compared with that of the designers. On a few occasions, the chain or logical framework that translates the defined problem into a series of concrete terms is valid in an intervention with objectives, instruments and operational actions. Design evaluation must analyse whether the context or environment of the intervention has been sufficiently considered and if the structure of those who politically promote the intervention has been suitably compared with the reference frameworks of citizens and stakeholders.

[5] Nominal Group Technique. See section on techniques and tools.

The evaluation of the aforementioned REINDUS programme (AEVAL, 2011) includes the following figure which groups all the stakeholders involved in the evaluated programme, their relationships, as well as the participating policies.

Figure 6. Involved stakeholders and relationships with the REINDUS Programme. Source: AEVAL (2011).



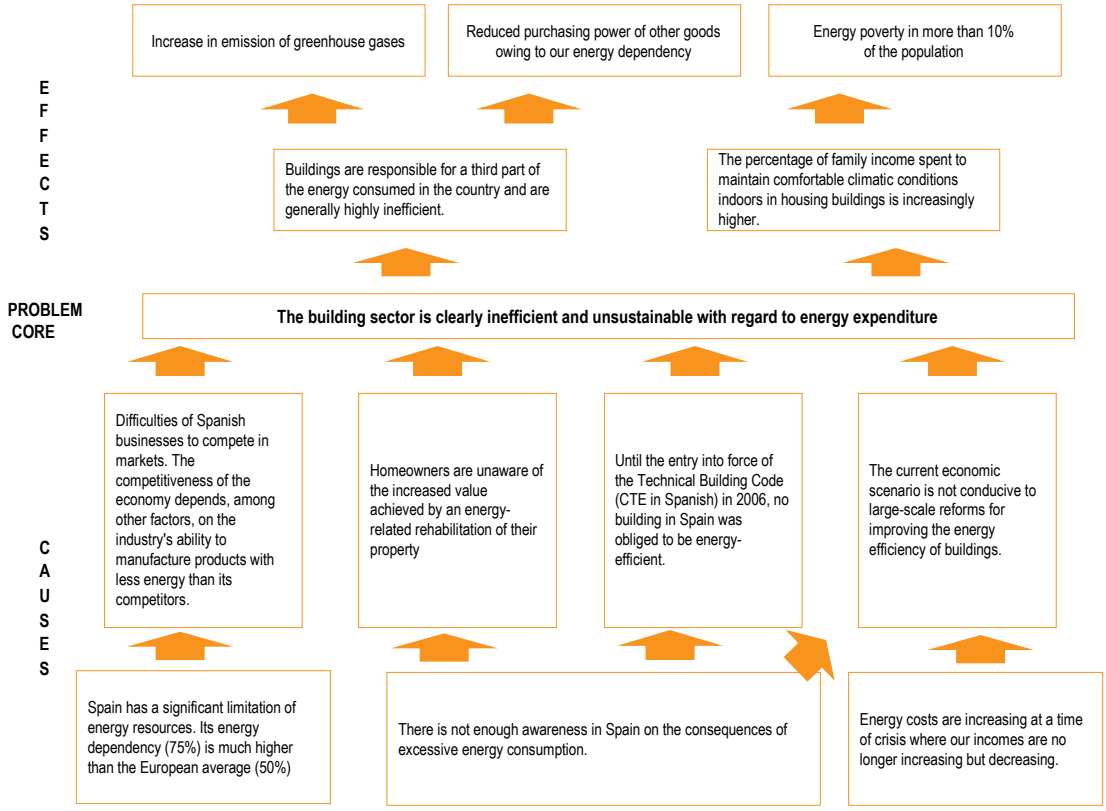
Problem Tree

Planning and design studies usually mention some useful tools for problem analysis. One such tool is the problem tree, which provides a quick view of the causes of a problem as well as its effects. The following stages must be taken into account when creating the graphic representation of the problem tree:

- Identifying the core problem of the intervention.
- Analysing and verifying the effects of the core problem. They are schematically represented above the core problem and may be classified into general and specific. If each first-level effect has other possible effects at the second-level, they are linked to each other..
- Establishing the relationship between the different effects caused by the core problem.
- Analysing the probable causes of the core problem. They are represented at a level below the core problem. It is possible to identify the cause of each negative effect of the core problem.

The problem tree is therefore linked to both the analysis of the problem and the relations of causality or causal logic, which shall be examined in other sections. It identifies the causes from a structural, causal, qualitative, and quantitative approach to the problem. It must not confuse causes with consequences or effects. On the basis of the problem tree, it is possible to create a tree of goals and areas of intervention, which is useful for analysing the intervention design itself. The following illustration displays the problem tree created for the Evaluation of the Plan for Energy Savings and Efficiency 2008-2012 (AEVAL, 2014).

Figure 7. Example of problem tree. Source: Evaluation of the Plan for Energy Savings and Efficiency 2008-2012. (AEVAL, 2014).



SWOT analysis and other techniques

Additionally, SWOT analyses allow us to have a holistic understanding of the problem as well as the intervention and its context. SWOT analysis is a technique that lists the Strengths, Weaknesses, Opportunities and Threats of a specific situation.

A multi-criterion, multi-objective approach (MOA) is a tool used to assess various possible solutions to a certain problem[6].

[6] For more detailed information on SWOT analysis, consult the section on tools and techniques in this Guidelines.

The choice of the public intervention. Alternatives and selection

All design evaluation must include the analysis of possible alternatives to solve the problem and the motives or circumstances behind the selection of each alternative. It is not about analysing the designed intervention, -which is performed in the concrete analysis of the design of the intervention- but about assessing it with regard to possible alternatives.

To be noted

In their analysis, the evaluator must always remember that establishing the agenda and suggesting alternatives are performed by different processes. Experts, academics, interest groups and public officials may be more important when creating alternatives whereas executive bodies, the media and public opinion may be decisive when it comes to establishing the agenda. While establishing the agenda is a more or less visible process, the selection of the alternative tends to be opaque.

This distvinction between the agenda and alternatives may be analytically useful and in spite of the fact that in a large part of the current literature on analysis public policies, agenda configuration refers simultaneously to both phenomena, and the distinction between agenda and alternatives is not clearly outlined.

Special attention must also be paid to the following aspects, some of which are common to the definition of the problem and its access to the public agenda:

- The importance and ***preferences of interest groups*** in selecting the alternative, adjusted to their definition and characterisation of the problem.

- The inclusion or ***explicit mention of different alternatives***. The policies whose consequences must be taken into consideration should be mentioned. It is important to include relevant alternatives. Beyond the set of possible alternatives, some alternatives are given greater consideration than others. The process of specifying alternatives reduces the alternative of conceivable alternatives that may be seriously considered or are feasible, given the context (political, economic-budgetary, lobbying groups, or purely political groups). The current policy must be included as a baseline, in order to estimate the improvement made by the other alternatives.

- ***Comparing the consequences*** of the alternatives with regard to relevance, cost and benefit, or the terms under which the best solution to the problem may be selected from the different alternatives. This analysis is performed for different scenarios, and the estimated costs and effects of each must be analysed, compensating or compromising between the different future scenarios. The possible trade-offs and consequences of selecting one alternative over others must be assessed, in terms of effectiveness.

The context of the intervention

The social, economic, institutional, regulatory, administrative, or other contexts or realities of the interventions are decisive in design evaluation. Public policies and programmes, and their interactions with different systems, including the political system, are produced in an environment and context, not in a vacuum or a controlled laboratory environment.

The context defines and conditions the design of a public policy and the will and motivations behind its execution. But it also defines its practice and results, that are but the effects of this intervention when interacting with the systems in all their scope (social, biological, and psychological, open or closed, etc.). After all, the different contexts act as the motors or the defining elements of the social problem, with regard to conditioning the intervention logic and as a motivator of the results of the implemented public policy.

The analysis of the context overlaps with the different components of the design evaluation. Therefore, it must be present in both the analysis of the problem and the selection of the alternative, the concrete design of the policy or programme or the theory of change. On the other hand, the context may have been modified at the time of performing the evaluation, therefore, it is important to be clear about the differences in context that may exist between the design stage of the intervention and its intervention.

When analysing the context of the design evaluation, the following contexts must be considered:

- The ***socio-economic context*** in which the intervention is produced. Both at the level of requirements and the availability of resources, a situation of decline, crisis or economic restrictions where requirements are higher or may appear with a higher intensity, is not the same as a positive economic cycle.

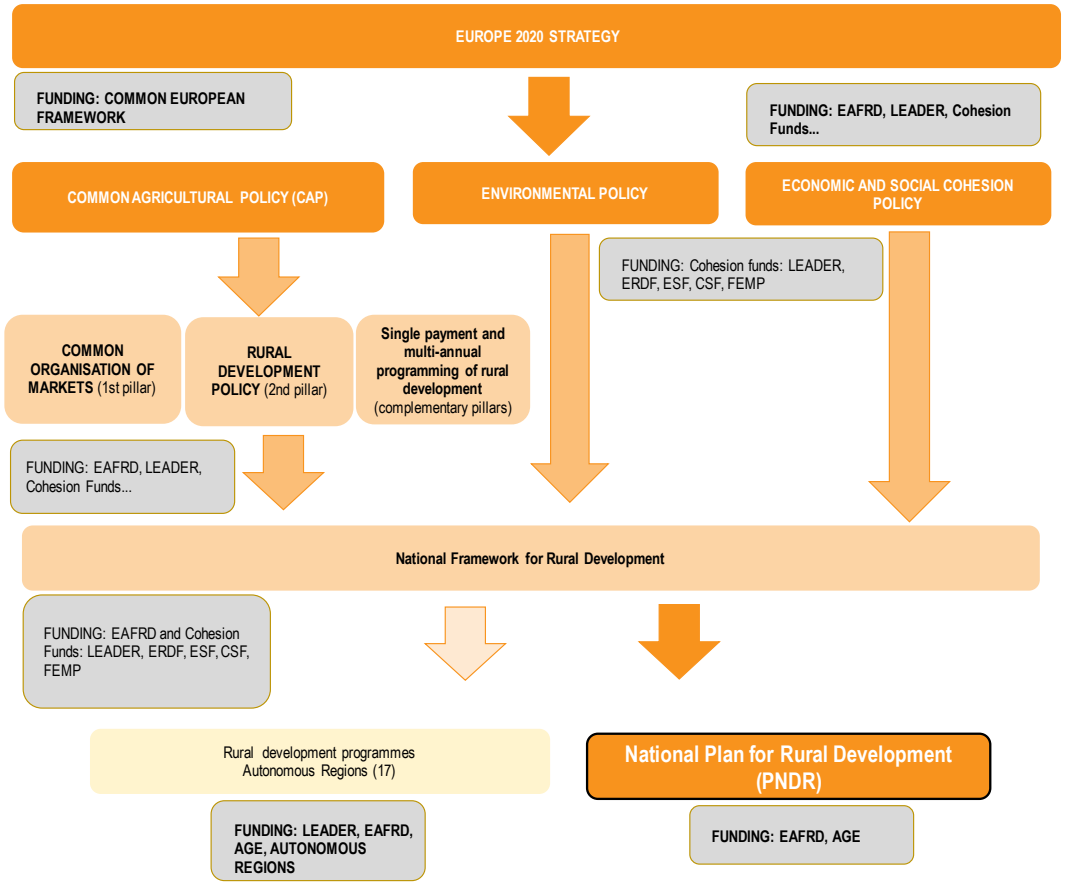
- The ***political context*** in which the problem, the alternatives and the design are defined. For example, a politically stable cycle, the end of a legislature, and the existence of a minority government are different.

- The ***legal and administrative context***: the set of standards and institutions that define a certain public policy field or sector where the intervention takes place. When outlining this aspect, the area-based framework and the levels of government that participate in the design of the intervention must be taken into account. This is especially important in countries such as Spain, that are highly decentralised and with a system of distribution of powers where concurrent and shared powers are prevalent. The analysis of the involved levels of government must range from supra-national bodies to local bodies, including bodies at the level of the autonomous regions and the State (AEVAL, 2015).

The Evaluation of the Strategic Plan for Road Safety 2005-2008 may be considered as an example of context-based analysis (AEVAL, 2009). It analysed the distribution of powers on traffic and the circulation of motor vehicles, which is the remit of the General State Administration Services, as well as basic legislation on other issues, some of them the responsibility of the Autonomous Regions.

It was also necessary to assess the implications of the European plan in this regard and the distribution of powers with regard to highway management, education, etc. of the Autonomous Regions and those of local city councils on traffic or circulation, among others.
It is important to seek alignment with other policies as especially in the case of regional policies, the evaluated policy may have a series of development tools of regional programmes with a specific source of funding. This was the case in the Evaluation of the Implementation of the National Programme for Rural Development 2014-2020 (AEVAL, 2017).

Figure 8. Alignment of the National Programme for Rural Development 2014-2020 and its sources of funding. Source: AEVAL, 2017.



Analysis of the formulation of the intervention

To analyse all the aspects of the problem, design evaluation may use any instrument or tool to respond to the questions posed by the evaluator. There is no single methodological tool that lets us analyse each question that is posed and mixed methods are preferable. The use of methodological triangulation which uses different methods that focus on the same social reality and verify whether the results are consistent, gives the analysis greater validity.

Both ***qualitative and quantitative techniques*** are recommended. The first includes discourse analysis, discussion groups, focus-groups, nominal group technique, and structured or open interviews. All of these techniques and group techniques in general, help us to make an in-depth study of different aspects.

Quantitative techniques enable the possibility of analysing and correctly defining the problem: its intensity or magnitude, identifying and quantifying the target population, etc.

Finally, documentary research and expert consultation is useful for learning about the birth of the intervention, its legal framework, the characterisation of the problem and underlying motivations, among other questions.

ANALYSIS OF THE INTERVENTION DESIGN

This section approaches the analysis of the design of the selected alternative, in terms of evaluation, the underlying causal theory and how this causal theory unfolds in a hierarchy of objectives and actions to respond to a detected problem. Along with the analysis of the problem and the analysis of the intervention logic, it is a part of the overall evaluation process of a public intervention.

Works on public policy planning and evaluation include various definitions of what is the causal theory of an intervention, which differ with regard to their scope and terminology rather than making a clear conceptual distinction. There are authors that refer to this causal theory as a chain of results, the theory of the programme or a theoretical model. They also mention the most utilised references of logical framework or model and the theory of change.

Nevertheless, neither the logical framework nor the theory of change may be considered synonyms rather they are complementary as they operate in different spheres of the design of an intervention. The logical framework belongs to the scope of action of the design and describes the logical sequence of how strategic decisions are taken in general and intermediate goals, activities, and resources to produce the awaited results and impacts. The Canadian Center of Excellence for Evaluation considers it a sequence of events and results (products, immediate results, intermediate results, and final results) expected to take place due to the intervention. A term to be found in the literature on this subject is short-cycle results chain. This guidelines refers to this logical framework as the ***theory of action or implementation***.

In contrast, the ***theory of change*** is located within the strategic sphere of design. Causal reflective reasoning explains the strategic options and outlines the premises of an intervention logic based on a desired change. Projects are drawn up and implemented within a logical framework to make the strategic decisions defined in the theory of change operational. Therefore, a sound theoretical framework incorporates a sound theory of change (Retolaza, 2018). The theory of change explains how an intervention in expected to achieve its results and explains its mechanisms in greater detail than the logical framework, as well as the assumptions, risks and context that supports or prevents the manifestation of the theory as observed results. It is also called a long-cycle results chain. The following figure shows the main distinctive and complementary characteristics of both elements.

Figure 9. Distinctive and complementary characteristics of the logical framework and the theory of change. Source: Author's own.

DIFFERENTIAL AND COMPLEMENTARY CHARACTERISTICS

LOGICAL FRAMEWORK	THEORY OF CHANGE
Operational Scope, Action of public intervention design	Strategic Scope of the design of public interventions
Descriptive Analysis of the sequence of actions	Explanatory Analysis of the Causal Logic Reasoning
Elements: Vision, Objectives, Outputs/Components, Activities and Resources	Elements: Assumptions/hypotheses, Strategies or mechanisms of each level, Conditioning Factors, Risks
It is a linear tool for the analytical and hierarchical structure of the elements or levels of the results chain	It is an explanatory tool that links the inter-connections in the roadmaps of change.
It focuses on the definition of the levels of action and their sequential alignment	It focuses on the mechanisms that explain intermediate and final changes, the roadmaps of change, and on the interactions or links of the results chain with the respective assumptions
At its core it deals with the levels of the results chain and the cause-effect relationships linking the results of each level and identifying the external assumptions that should be considered but without analysing the underlying assumptions that define the changes	At its core it deals with the underlying explanatory assumptions of the cause-effect relations and the strategies defined to achieve the changes, as well as the assessment of the risks associated with internal and external conditioning factors
The Assumptions (hypothesis) are directed towards identifying the questions related to the environment and therefore external ones beyond our control that affect the intervention	Assumptions refer primarily to the reasons that explain each change at the different levels taking into consideration both internal and external questions
The result is the narrative of the articulation of the logical sequence of activities, outputs, results and impacts that allow us to achieve the objectives and the vision of the intervention	The result is the underlying explanatory theory, a reflective and reasoned thinking of the changes generated by the intervention, of how they are produced at different levels of the intervention until the final desired change
Graphic representation: Logical Framework Matrix or Results Chain	Graphic representation: Diagram of inter-relations, nodes or networks and Assumptions Table

Theory of action or implementation of an intervention

The theory of action or implementation is one of the components of design analysis that is also referred to by different authors as the logical model. Regardless of its name, it refers to the sequential structure of the intervention with regard to goals, activities and resources that lead to the outputs that in turn generate the results and impacts awaited from a public intervention.

They are basically descriptions of the implementation of the different elements that constitute the intervention and appear throughout the cycle of action, as all interventions may undergo changes during their execution due to (external or internal) unforeseen circumstances or due to the detection of defects or deviations in execution.

The theory of action allows us to comprehend and assess the structure of a public intervention and its implementation at any stage, either planning, monitoring or evaluation. In the first stage, a sequential layout of the theory of the action allows us to better define and programme the activities and resources required to execute the plan or programme. The monitoring stage allows us to identify and correct deviations in implementation. Finally, in the evaluation stage, the theory of action allows us to better understand the intervention, assess its implementation, identify the evidence of its design coherence and provide explanatory information on the results obtained.

In design evaluation, the first stage of the analysis of the theory of action is to build or validate the hierarchy of objectives. All public interventions must have a general objective that splits up sequentially into strategic and operational objectives, and which must be related to the measures that seek to achieve said general objective by means of a series of resources and activities.

To be noted

Listing the different elements that constitute the intervention (strategic goals, operational goals, measures, awaited results and impacts) gives us a better understanding of the intervention and helps us to draw us the evaluation questions, mainly in design and implementation evaluations.

Nevertheless, the hierarchy of objectives is not always clear or explicit in the intervention, or it may be inverted at some level of the objectives. It is therefore necessary to build or validate it by means of different techniques with the participation of key actors. The importance of identifying the structure of objectives lies in its importance when analysing the coherence of the design with the problem to be solved, with other interventions or with the measures or activities to be defined.

Objective Tree

One of the most frequently-used tools for graphically arranging these analyses is to create an ***objective tree***. The objective tree is a methodological procedure that identifies and classifies objectives according to their importance and displays the means-goals relationships in a diagram.

The objective tree may be created independently or as a complement to the problem tree discussed in the section on the analysis of the problem and "consists of converting the negative states of the problem tree into solutions, expressed as positive states.

As a matter of fact, all of these positive states are objectives and are presented in a diagram of objectives which displays the hierarchy of the means and the goals" (CEPAL, 2005) Here, the core problem would be the main objective and the effects would become the goals of the intervention.

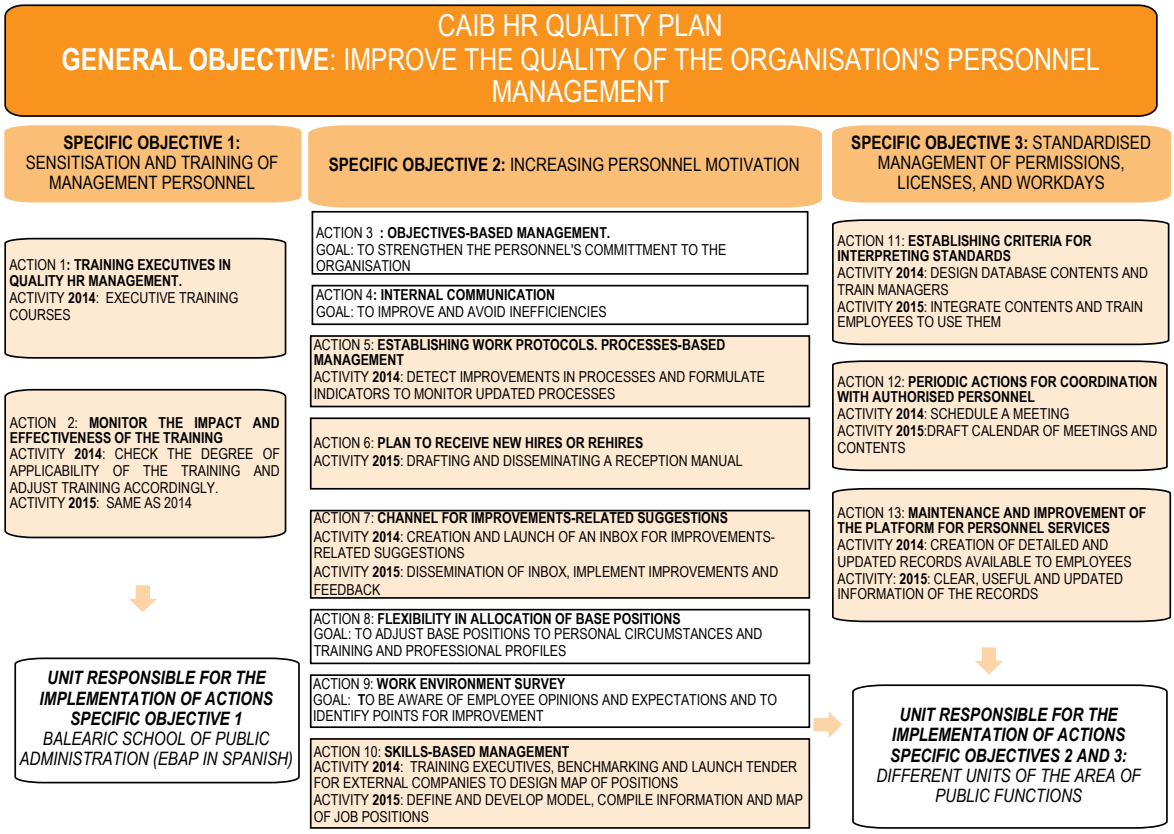
- The objective tree is built in the following manner:
- The problem is converted into a desirable positive state.
- Analyse the causes of the problem, formulate them as positive states and adjust them to the general objective.

Analyse the means-objectives-goals relationships to ensure the consistency of analysis.
For example, in the "Evaluation of the Human Resources Quality Plan of the general services of the Administration of the Autonomous Regions of the Balearic Islands" (AEVAL, 2015), the general objective was to improve the quality of the organisation's personnel management. To achieve this, three specific objectives were proposed: the sensitisation and training of management personnel, increasing personnel motivation and the standardised management

of permissions, licenses, and workdays.

These objectives in turn break down into different actions that are the responsibility of different units. Once the analytical hierarchy of objectives has been identified, the following step is to relate them to the activities, the measures designed to achieve them, with the allocated resources and with the offered products and services.

Figure 10. Objective tree. Source: Evaluation of the Human Resources Quality Plan of the general services of the Administration of the Autonomous Regions of the Balearic Islands (AEVAL, 2015).



The logic is the following: using a series of inputs (human and financial resources, etc.) a series of activities are performed that provide results on the operational objectives. This in turn leads to changes in the affected variables (strategic objectives) and the socio-economic environment of the intervention (general objective).

In the theory of action, the strategy must be validated and assessed to ascertain whether the actions are consistent with the objectives, by means of an orderly strategy that, in the words of Martinic "*are the means or roads that, in the opinion of the project authors, are the most suitable and effective to reach the proposed goals and changes*." (Martinic, 1996)

Different tools for graphical or schematic representation of the theory of action are used, more or less complex models according to the type of intervention, often called logical models.

Although there is a large variety (realist matrix, triple column, Tiny Tools results chain, etc.) the most frequently used ones are the results chain and the logical framework matrix[7].

Results Chain

The ***results chain*** consists of a linear representation with boxes/squares of objectives, activities, products, results and impacts. It is a useful tool for representing simple interventions.

Figure 11. Results chain. Source: Author's own.

For more complex interventions, the ***logical framework*** matrix is more suitable than the results chain.

Logical Framework Matrix

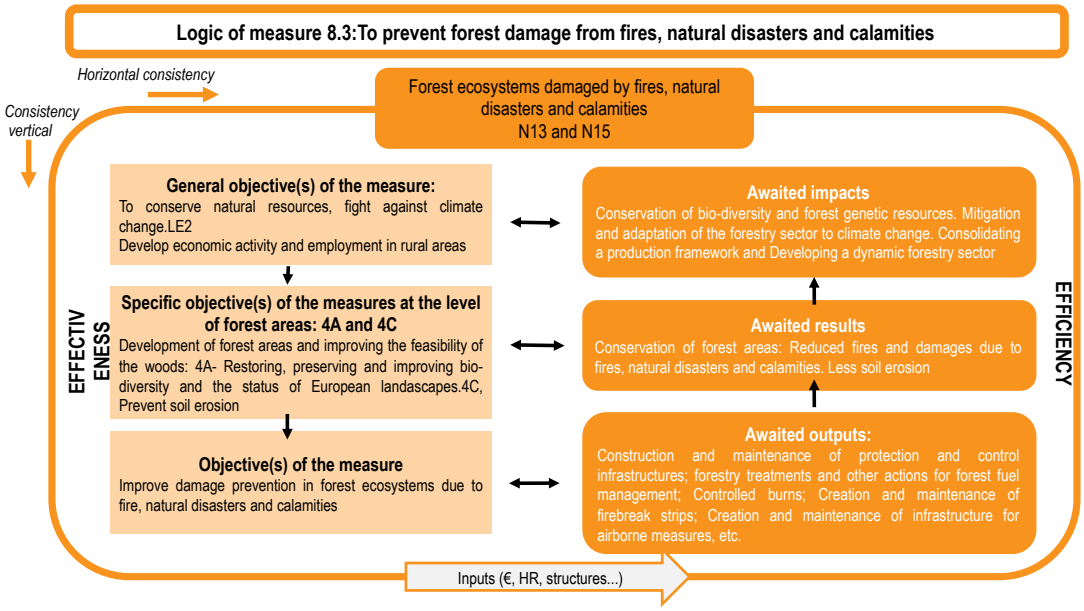
It is a matrix representation that provides a horizontal and vertical reading to obtain the causal relationships and the narrative of the logic. This technique has limitations, as it does not reflect all the interactions. The narrative derived from the logical framework matrix is the basis for the theory of change which must be supplemented with the analysis of the assumptions and conditioning elements (internal and external factors to be taken into account for achieving successive assumptions and finally for achieving the vision).

On the basis of the logical framework matrix, we may define the timeline of the activities and the indicators of each for their monitoring. It is a planning but also a management instrument that allows the persons or bodies in charge to ensure the quality of the planning and the implementation of the intervention.
In evaluation, it gives us a better understanding of the intervention and the design of the evaluation as it steers the creation of the evaluation questions mainly in design and implementation evaluations.

The logical framework matrix is the tool to assess the implementation of the policy, plan, or programme in an evaluation. Below is an example of the graphical representation of a logical framework matrix.

[7] *Tiny Tools* results chain: mapping potential positive and negative impacts of an intervention. Logical framework: the design, execution and evaluation of projects taking into account the relationship between the available resources, the planned activities and the desired changes or results. Realist matrix: focusing on one of the steps in a results chain and subsequently identifying the mechanism involved in producing the results and the contexts in which this mechanism operates. Results chain (also known as a 'pipe model'): showing a programme as a series of input boxes -> activities -> outputs -> results -> impacts. Triple column: it shows a hierarchy of the results in the central column.

Figure 12. Logical framework matrix. Source: Intermediate Evaluation of the National Plan for Rural Development (AEVAL, 2017).



Theory of change or causal logic of an intervention

As we have seen., any public intervention must be based on a ***causal theory***, that is to say, on a set of assumptions, conditioning factors and external factors that describe how and how the programme is meant to function and obtain certain awaited results.

The theory of change is defined as how causal reflective reasoning explains the strategic options and outlines the premises of an intervention logic based on a desired change. Projects are drawn up and implemented within a logical framework to make the strategic decisions defined in the theory of change operational.

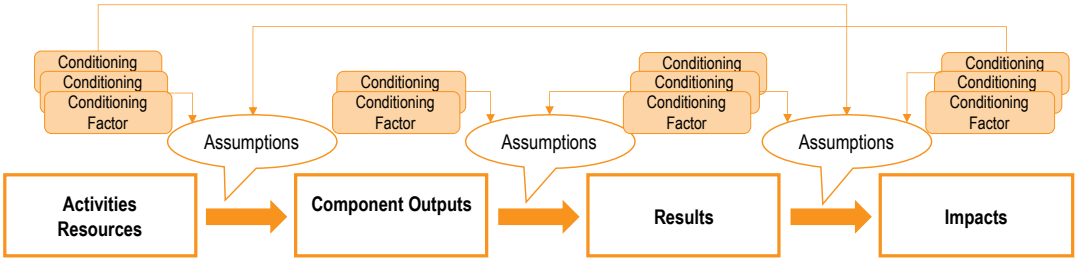
The theory of change -also called the theory, hypothesis or causal logic of the intervention- is inherent to the design of any public intervention, is located within the ***strategic sphere of the design*** and generally appears either explicitly or implicitly in the formulation and design of public policies.

It refers to how the intervention seeks to generate the required changes at different stages or phases to achieve the intermediate results and the expected final transformation. It defines the strategies to be followed, taking into consideration the risks and factors that may influence the achievement of these changes, as well as the conditioning factors that are required to achieve them.

Graphical representation of the theory of change

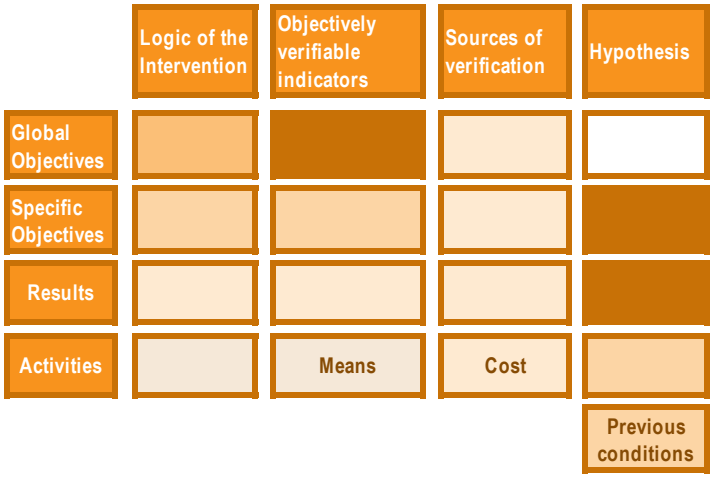
The theory of change may be conceived of as the story of what needs to happen in the arrows that link the boxes in a traditional logical model of a results chain, as shown in the following figure.

Figure 13. Diagram of the theory of change as a results chain. Source: Author's own.



The theory of change may also be expressed as the causal relationships between the results and the different levels within the logical framework matrix. Some logical framework models, such as the logical framework of the European Commission, include a column of assumptions or hypotheses to include the analysis of the theory of change of the interventions, as may be observed in the following figure.

Figure 14. Logical framework model. Source: European Commission (2001).



This option is simple and easy to apply and understand. However, it is limited when it comes to complex interventions as it does not display the multiple causal or conditioning relationships that intervene in the assumptions of the different stages or intermediate phases. For more complex interventions, it is recommended to use an exclusive graphic representation of assumptions and conditioning factors that provide a visualisation of their inter-connected nature for better comprehending the intervention. Representations that use network diagrams or a specific table of assumptions and conditioning factors may be used.

To be noted

Depending on the different interpretations of the key stakeholders, when performing an evaluation, the assumptions, risks, and approaches shall define different theories of change. The evaluator must identify and analyse these interpretations in order to build an overall theory of change of the intervention or to validate that which is enshrined in the documents on its formalisation.

As Vogel[8] states, the theory of change is not a single exercise to design (or evaluate) an initiative, but it involves continuous learning and an adaptable management cycle. It is a process where it is necessary to have all the key stakeholders of the intervention in order to rebuild the hypothesis in the intervention formulation and design stages and the circumstances that have played a role in the implementation and results stage.

The design evaluation attempts to identify the underlying bases of the internal cause-effect logic of the public programme or policy and to question and validate the relationships between the immediate outputs of the intervention and the medium-term and long-term results.

The analysis of the theory of change on the design of the intervention has a two-fold goal:

- On one hand, it allows us to assess the consistency and quality of the ***underlying theory*** of the intervention at the level of the defined strategies, i.e., how to solve the requirements or problems of the target population, and the contribution of the implemented measures or activities with regard to the distribution of the input resources, and outputs or services aimed at solving the problem.

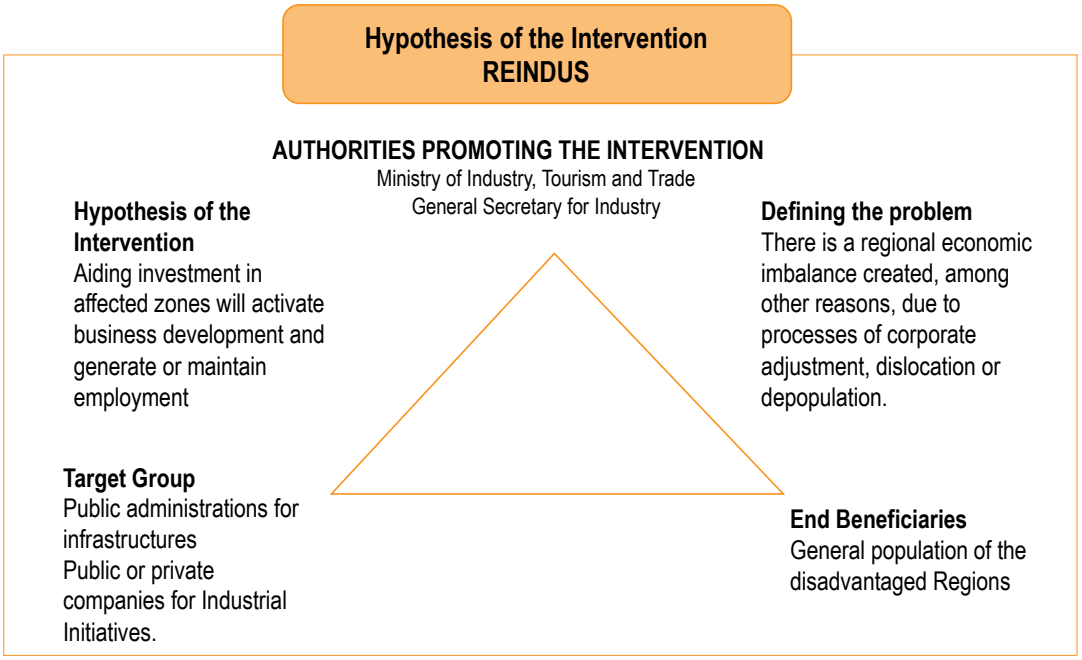
- Additionally, in the case of an ex-post evaluation, it helps to identify whether there have been ***deviations in the intervention*** as a result of the design mechanisms themselves, the process of implementation, the changes to the context, the effect of incentives and disincentives, etc.

When evaluating public policies, the analysis of the theory of change is based on the narrative of the intervention logic to subsequently analyse the required hypotheses or assumptions that must be present in the roadmap of changes at each level of objectives to ensure the success of the intervention.

Narrative of the intervention logic

The concept of the narrative of the intervention logic refers to the alignment of the activities, outputs, or components, results and impacts of the intervention to be evaluated. The narrative corresponds to the initial hypothesis of the intervention, that is to say, its theory and the strategies for achieving the intermediate and final changes that are sought. Occasionally, the explicit hypothesis may generally be found in the documents formalising an intervention. In these cases, it is the responsibility of the evaluators to validate the hypothesis at the time of the evaluation.

Figure 15. Hypothesis of the intervention of the REINDUS Programme based on the Means Guidelines. Source: AEVAL (2011).

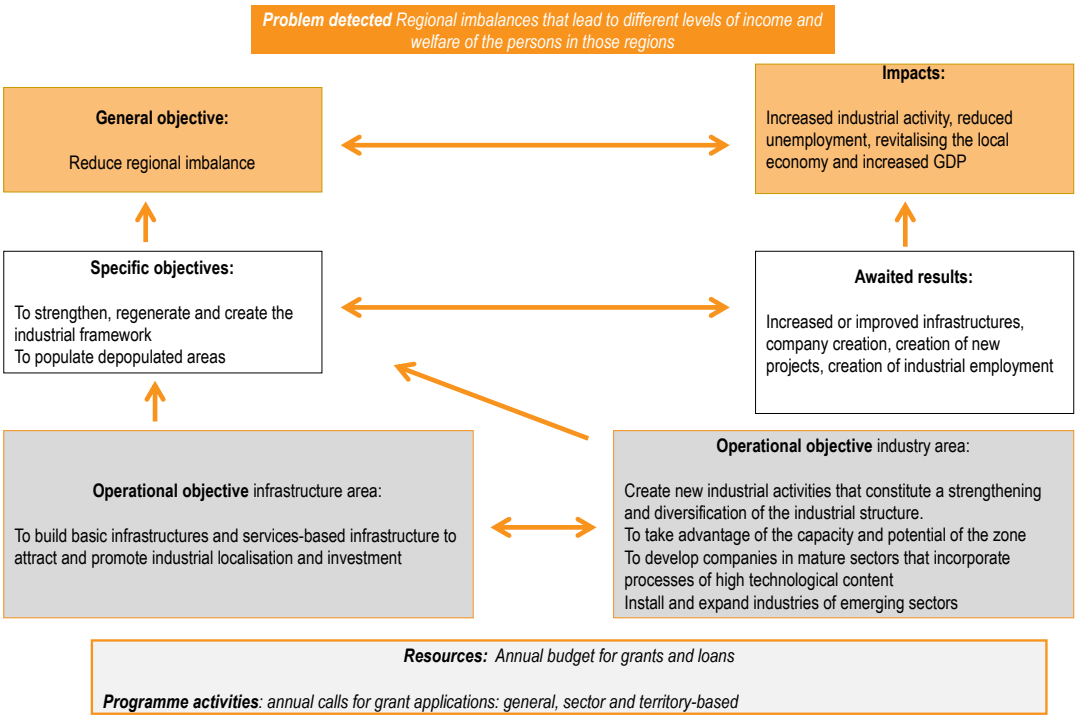


On other occasions, this hypothesis is not explicit or clear and therefore, it is necessary to reconstruct it at the time of the evaluation from documents and the involvement of the persons or bodies in charge, the managers, and the key stakeholders. The following figure presents the hypothesis of the REINDUS programme (AEVAL, 2011): the awarding of grants for investment would reduce existing regional imbalances in disadvantaged zones due to deindustrialisation processes.

The hypothesis is completed with the hierarchical breakdown of the general objectives into one or many specific or operational objectives, that is to say, with the analytical structure and the strategies defined to achieve the objectives, which are the strategies that drive the theory of change. The strategies may be defined as the roadmaps of the change or the set of operations grouped into one or more areas of action of the intervention to achieve its intermediate goals.

In the case of REINDUS, action strategies are focused on two areas, infrastructure, and industry, as may be observed in the following figure.

Figure 16. Theoretical model of the intervention logic of the REINDUS Programme Source: AEVAL (2011).



When analysing the theory of change of a design evaluation, the evaluators must respond to the following questions: *What are the strategic areas on which the action of the intervention centres and how do they act?* What are the interactions between them? It must be considered that the answer to these questions varies according to the different points of view that the key stakeholders have on how to achieve the goals in different aspects or dimensions to achieve the final desired change. Each key stakeholder will thus provide a specific vision of how the intervention generates changes, providing the evaluator the most complete information possible to reconstruct or validate this part of the intervention design.

The techniques used to obtain all the required information to create, reconstruct or validate the narrative, mainly documentary research and qualitative techniques including interviews, discussion groups or Nominal Group Technique. Complex interventions may require surveys to collect the points of view of a majority of the stakeholders, for example in strategic plans with a cross-cutting scope that reference powers shared between the General State Administration Services, the Autonomous Regions and Local Bodies, where the responsibility is divided among many managers.

Narrative of the logic of the intervention

A useful tool in evaluation tasks is the simple and concise graphic representation of the results of the applied techniques that help us in the analysis. There is no single form of representation; a hierarchy-based plan of analysis may be used such as the aforementioned triangle of intervention (European Commission. EuropeAid - Cooperation Office, 2001), results chain, theoretical framework or logical framework model; and each evaluator shall select the most suitable technique in each case. In the figure depicting the "*Logical Framework Model*" of the European Commission, the column headed "intervention logic" lists the narrative of said logic.

Analysing assumptions

One of the elements to be considered in the theory of change are the assumptions, defined as the sufficient circumstances, conditions or events that must be present for the awaited change to take place at each level of the objectives. They are the reasoning or theoretical-predictive arguments, including exogenous situations (external conditions that are beyond the managers' control) that must be present for achieving the results at each intermediate stage and at the end of the intervention.

Weiss popularised the term "theory of change" as a way to describe this set of assumptions that explain the intermediate steps that lead to the long-term target of interest, as well as the connections between the programme activities and the consequent results at each step along the way. Figure 13 "Diagram of the theory of change as a results chain" displays these assumptions.

Some authors consider assumptions to solely constitute of the circumstances that are necessary and external to the intervention, beyond the control of the persons or bodies in charge of it. Others give it a wider meaning and include the belief systems of the different involved parties or stakeholders, either internal or external to the intervention, on how the change will take place (Retolaza, 2018)[9]. Regardless, what is of interest in design evaluation is to identify and select those assumptions that are relevant and critical for producing the change at different levels of objectives.

The questions associated with these assumptions are: *What conditions or circumstances (short, medium, and long-term, simultaneous or not) are required to achieve the desired change?* Identifying these conditions may be a complex process. Although assumptions are coded positively, they are in fact risks that may compromise the execution and results of the intervention. For this reason, the methodology of risk management is applied to the analysis in order to predict their appearance or to reduce their impact if they do materialise.

Analysing assumptions enriches the design if it is performed at the planning stage of the intervention and if it is performed at the time of the evaluation, it gives a better understanding and comprehension of the latter. Design evaluation uses it to analyse the coherence and effectiveness of the design, i.e., to analyse the degree to which the design is in line with the objectives and results to be achieved (both short and long-range[10]). It is also used to evaluate the implementation of the intervention and the system for monitoring and controlling the indicators associated with the implementation risks. It can also offer pertinent information to analyse complementarity.

[9] *Following Iñigo Retolaza, managing assumptions focuses on explaining why the intervention will make an effective contribution to the desired change. By shedding light on the underlying assumptions, it seeks to explain how interventions and activities will make a significant contribution to achieving certain intermediate and final results.*
[10] *Long-range results may also be termed impacts.*

The evaluator identifies and analyses the assumptions with the participation of the key stakeholders (with regard to their significance and probability) to select those that are critical or present significant levels of risk. With regard to importance, only important assumptions are collected for each desired change, that is to say, assumptions that respond to the question: *"If this assumption were absent, could the plan or objective be executed in order to obtain the result or output, or to perform the designed activity?"*

Once the important assumptions of each change have been identified, the probability of occurrence of each is defined (the certainty of the appearance of this assumption). There are two possible situations:

- Assumptions with a high probability of occurrence, with a high certainty of appearance. This case deems it a strong assumption to account for the desired change in the element or stage to which it is related. It also means that there are no conditioning factors that compromise its appearance, or they are of little relevance, and therefore do not constitute a risk to the success of the intervention. These assumptions do not require mechanisms for monitoring and indicators for their control.

Additionally, we must also take into account that interventions are dynamic processes influenced by the context in which they take place and the circumstances of their implementation, which may lead to a variation in the initial assumptions that were considered in the planning. Therefore it is necessary to validate the initial assumptions and redefine them according to the current situation or include new ones if necessary at the time of the evaluation, in order to enhance the coherence of the intervention and to explain the results obtained. It serves as a support to the conclusions and recommendations that are generated at the end of the evaluation process. The guiding question would be *"To what degree have the initial assumptions changed, and what are the causes and results of these changes?"*

Conditioning factors that influence assumptions

The assumptions present in the roadmap of change at each level of the objectives may be influenced by various external conditioning factors (social, economic, regulatory, political, etc.) that are outside of our framework of action. Therefore, they constitute threats or vulnerabilities that cannot be controlled and may promote or obstruct the presence of these assumptions. They respond to the question: "*What conditions and actors promote or present obstacles to the desired roadmap of change?*"

Similar to the assumptions, conditioning factors are analysed according to the level of risk posed by the probability of their materialisation and their influence on or relevance to the element under analysis. That is to say, it measures the influence of these factors in not producing the hypothesis or assumption that determines the change in a stage or phase.

Conditioning factors with a high degree of probability and influence are the ones with the highest levels of risk and therefore there must be mechanisms to monitor and control them during the implementation of the intervention. Conversely, conditioning factors with a low or very low probability of occurrence or influence are less or not responsible for the failure of the assumption or hypothesis and therefore, do not require targeted mechanisms for control.

Assumptions are identified or defined on the basis of documentary research and this is complemented with the participation of the key stakeholders by means of qualitative techniques such as in-depth interviews, creating a SWOT analysis or a Nominal Group Technique.

The participation of multiple stakeholders enriches the analysis by including different points of view or interpretations of how the intervention can achieve the final objective. Or to put it in other words, different and particular theories of change are identified as each participant has a different narrative on how the changes are produced at each level of the intervention, passing on to the next participant until we finally arrive at the awaited impacts.
The evaluator must identify the theory of the overall change of the intervention.

Assumptions Table

As a support tool for the analysis, different methods for visually representing the assumptions are used such as the assumptions table, the logical framework matrix, or network diagrams. Here we have opted for the assumptions table.

Figure 17. Table for the analysis of assumptions and conditioning factors of the theory of change. Source: Author's own.

Vertical narrative Hierarchy of Objectives	ASSUMPTIONS/HYPOTHESIS		CONDITIONING FACTORS			
	DEFINITION	PROBABILITY OF OCCURENCE %	IDENTIFICATION	PROBABILITY OF MATERIALISATION %	INFLUENCE	LEVEL OF RISK
					Low-Medium-High-Very High	Null-Low-Medium-High
GENERAL OBJECTIVE/AIM						
SPECIFIC OBJECTIVES OBJ 1 OBJ 2....						
RESULTS R1 R2....						
COMPONENTS/OUTPUTS C1 C2....						
ACTIVITIES A1 A2....						

The assumptions table of the theory of change allows us to analyse the different assumptions and conditioning factors by assigning a level of risk according to their probability of appearance and the influence of this conditioning factor in achieving the assumption.

The table helps to reflect and take decisions on the action that the persons or bodies in charge must take with regard to the involved risks. The risks may be eliminated, transferred, mitigated, or assumed. The choice of action to be taken depends on the difficulty and cost of countering the risk and its relevance.

The evaluation analyses the existence of these risks and the actions or decisions taken by the persons or bodies in charge, the managers of the intervention to control or eliminate them. The question to be asked at the time of the evaluation is: *"Have the persons or bodies in charge or the managers of the intervention identified the risks and taken decisions, once the level of risk of the conditioning factors has been ascertained, in order to ensure the success of the intervention?"*

The logical framework matrix includes a column with the hypothesis or assumptions of the intervention. It uses a clear and concise tool to demonstrate the theory of change in simple interventions, but is of less use in complex interventions owing to the rigidity of its design, when assumptions and conditioning factors may be related with various intermediate elements of the intervention logic but cannot be easily visualised with this representation.

To be noted

Normally, in ***ex-post evaluations,*** *the assumptions table is created but not presented in the evaluation report. Instead the results are included in the analysis of the criteria of suitability, coherence, and implementation. Other cases do not have a matrix of the theory of change but these assumptions and factors are analysed on the basis of the results of the techniques used (interviews, group discussions, surveys on the design and implementation, etc.), with the evaluation report including the conclusions and recommendations for their monitoring and control.*

An example of risk analysis in evaluation is the one performed by AEVAL in the Evaluation of the National Plan for Transition to Digital Terrestrial Television (AEVAL, 2009). The objective of the plan was to define the directives for the progressive analogue switch-off in Spain by technical areas, through 90 technical transition projects in the entire national territory. The recommendation on identified risks that might make it difficult to achieve the programmed objectives and thus the effectiveness of the National Plan for Transition to DTT includes the conclusions of the application of the methodology for identification, risk assessment and existing preventive and corrective measures in the intervention.

Example of risk analysis. Evaluation of the National Plan for Transition to Digital Terrestrial Television (AEVAL, 2009). Recommendations with regard to the NTP-DTT

The main risk factor for the fulfilment of the NTP-DTT are the delays in the actions to extend coverage that are the remit of the Autonomous Regions. The reason for said risk lie in the lack of time to perform these actions due to the delays in establishing agreements between the central and regional governments on how to finance the coverage extensions, as well as the publication and awarding of the required public tenders. For this reason, it is recommended to review the current management model of public funds and their implementation, as well as to analyse alternative models that will allow us to expedite the schedules and reduce the impact of the coverage extensions on the overall development of the NTP-DTT.

One of the elements not foreseen by the NTP-DTT is the phenomenon of the gradual analogue switch-off in a Transition Project. This situation, known as a "sliding switch-off" and classified by the present study as a technical risk of low relevance owing to the opinions of the key stakeholders, currently qualifies as a failure according to the formulation of the NTP-DTT. Therefore, it is recommended that such switch-offs be classified as natural, reasonable, and necessary technical elements that are considered as successes in terms of the evaluation.

Data on the implementation of the transition to satellite in areas of DTT coverage that do not have access to the signal and have proceeded to switch-off analogue signals indicate by their extreme scarcity, that there is a need to perform an in-depth analysis of the current situation of the process and its evolution in Stages II and III.
A standardisation of the logos and brands that unequivocally denote the capacities of the receiving equipment is recommended, as well as providing citizens with clear information and guidelineslines of DTT products.

The implementation of mechanisms for a cascade model of training is recommended which provides local agents (consumer associations) with sufficient knowledge to channel incidences and solve most basic problems that may emerge in the transition process.

Synergy and complementarity of objectives

A final aspect to be analysed in the design of an intervention is the possible existence of synergies, complementarities, or secondary contributions. It allows us to check and assess the internal or external coherence of the intervention.

The achievement of the general or strategic objectives in many public interventions may be affected by the positive or negative influence of certain objectives over others. We may then state there are synergies, if they are positive, or antagonisms if the influence is negative. These influences arise at the level of the objectives but also at the level of the measures, in complex interventions that are considered as such provided they contribute to objectives other than that for which they were designed. The analysis allows the evaluator to identify the snowball or carry-over effect of the objectives or the level of influence and sensitivity of certain measures over others, and consequently to assess the effects of the interactions in the results obtained, given that an objective cannot be achieved by the results of the problems of implementing other carry-over objectives.

The contributions made by the intervention to achieving the objectives of other interventions are usually deemed complementary or secondary contributions according to the European Commission's terminology on structural funds. This generation of synergies is especially relevant when it takes place between the different levels of government in a highly decentralised nation such as Spain, which places an added layer of complexity on the complementarity and external coherence of the interventions to be evaluated.

Identifying and evaluating synergies requires the participation of the stakeholders through qualitative techniques and the application of tools to analyse the results obtained.

Synergy Matrix

The ***synergy matrix*** is usually utilised to assess synergies. The matrix used to evaluate the National Programme for Rural Development 2014-2020 (AEVAL, 2017) shows how, at the level of the sub-measures, different levels of effect may be assigned between them, in order to determine the level of influence on the achievement of specific objectives assigned to each.

Figure 18. Scores assigned by the managers of the measures to assess the synergy between the sub-measures of the National Programme for Rural Development. Source: AEVAL (2017).

	SUB-MEASURES														
	M1.1.	M1.2	M4.2	M4.3.1.	M4.3.2	M7.8	M8.3	M8.4	M9.1	M15.2	M16.1	M16.1+16.2.	M16.2	M16.5	M16.6
M1.1.		3	3	0	0	0	0	0	0	0	0	0	0	3	3
M1.2	3		2	0	0	0	0	0	0	0	0	0	0	2	2
M4.2	2	2		0	0	0	0	0	0	0	0	0	0	3	3
M4.3.1.	0	0	0		0	0	0	0	0	0	0	0	0	0	0
M4.3.2	0	0	0	0		0	0	0	0	0	0	0	0	0	0
M7.8	0	0	0	0	0		0	0	0	0	0	0	0	0	0
M8.3	0	1	0	0	0	0		0	0	0	0	0	0	0	0
M8.4	0	1	0	0	0	0	0		0	0	0	0	0	0	0
M9.1	0	0	0	0	0	0	0	0		0	0	0	0	0	0
M15.2	0	0	0	0	0	0	0	0	0		0	0	0	0	0
M16.1	1	2	3	1	1	0	1	1	2	2		5	4	4	4
M16.1+16.2.	1	2	3	2	2	0	1	1	4	2	3		5	3	3
M16.2	0	0	0	0	0	0	0	0	0	0	0	0		0	0
M16.5	2	2	4	0	0	0	0	0	0	0	0	0	0		2
M16.6	2	2	4	0	0	0	0	0	0	0	0	0	0	2	

M1.1.	Aid for professional training and skill acquisition
M1.2	Aid for demonstration activities and information actions
M4.2	Aid for investment in transformation/marketing and/or development of agricultural products
M4.3.	Aid for investments in infrastructures related to the development, modernisation or adaptation of agriculture and forestry
M7.8	Others
M8.3	Aid for preventing damage caused to forests by fires, natural disasters and calamities
M8.4	Aid for repairing serious damage caused to forests by fires, natural disasters and calamities
M9	Aid for creating groups and organisations of producers in the agricultural and forestry sectors
M15.2	Aid for the conservation and promotion of forest genetic resources
M16.1	Aid for the creation and functioning of EIP operational groups for agricultural productivity and sustainability
M16.1+16.2.	Aid for EIP operational group projects for agricultural productivity and sustainability
M16.2	Aid for pilot projects and the development of new products, practices, processes and technologies
M16.3	Cooperation between small operators to organise common work processes and share facilities and resources, such as the development or marketing of tourism
M16.4.	Aid for horizontal and vertical cooperation between agents of the supply chain in order to implement and develop short supply chains and local markets, and for promotional activities within a local context linked to the development of short supply chains and local markets
M16.5	Aid for joint actions performed to mitigate climate change and adaptation, and for joint approaches with regard to on-going environmental projects and practices
M16.6	Aid for cooperation between agents of the supply chain for the sustainable supply of biomass for food preparation and energy production and industrial processes

The results of this table display a low level of influence, therefore, the evaluation concluded that the measures were considerably independent and there was insufficient complementarity between them for synergies to be produced by their interaction.

Tools for design analysis

There are various tools that are useful for analysing the design of the intervention, both for the theory of action and the theory of change.

Apart from reviewing the key documents of the intervention, other qualitative techniques such as ***semi-structured interviews***, ***discussion groups*** and ***nominal group techniques*** may be used to reconstruct the logic of the intervention and within it, the underlying theory of change. In any case, it is important to have the participation of all the stakeholders to obtain all the different approaches, points of view and different interests that allow evaluators to reconstruct the theory of change or to assess if it is explained in the document formalising the plan or programme to be evaluated.

In the ***stakeholders interviews***, questions are posed on the intervention, on its design and structure, and how the results are obtained.

To identify the assumptions in the evaluation, in addition to the aforementioned techniques , the ***SWOT analysis technique*** with key stakeholders of the intervention. The lessons that are identified in interviews with the persons or bodies in charge and the managers of other similar or earlier processes are also a source of information when defining conditioning factors.

EVALUATION CRITERIA AND QUESTIONS

The ***evaluation matrix*** is the basic tool that steers any process of evaluation. It consists of different components essentially evaluation questions and criteria that are the different perspectives, dimensions or approaches to be explained. It also includes the indicators, sources, tools and techniques, and other data to provide solidity and credibility to the conclusions and recommendations that are based on the responses to the questions of the included criteria.

Evaluation criteria in design evaluation

An interesting definition of criteria which has its origin in the field of education is that they are "*the principles, standards or ideas of assessment based on which a judgement regarding the evaluated object is made*" (García Sánchez, 2010).

Based on this definition, the criteria for this Guidelines may be defined as the different points of view or approaches that are made on the target of the evaluation. They are conditions, rules, principles, standards or ideas of assessment that make it possible to issue a judgement on that which is evaluated.

The criteria provide benchmarks that make it possible to obtain useful knowledge to make assessments. Additionally, they allow *questions to have a structure -as a matter of fact, they act as guideliness for their formulation- and the focus of the evaluation, covering the field or dimensions of a public programme or policy to be evaluated. Indeed, the questions themselves usually belong to different families of criteria"* (AEVAL, 2015).

From the point of view of design evaluation, the benchmark criteria are the following: suitability, relevance, coverage, internal coherence, external coherence and complementarity.

Suitability

Suitability is one of the principal criteria in design evaluation, involving the degree to which the set of measures that constitute the intervention is directed towards solving the existing problem or requirement within the context in which the problem is generated. It may also be defined as the degree to which the objectives of an intervention are consistent with the existing requirements. Thus, if the requirements do not exist or are not those that were formally identified, the intervention is not suitable as it is not focused on the problems that motivated it.

Relevance

The criterion of relevance refers to the degree of importance of the role of the measures within the wider framework of the policy. The interest lies in the set of measures as a tool of the intervention. Therefore, given the scale and breadth of a problem, or when it is necessary to mobilise important economic-budgetary resources, an intervention will not be deemed relevant if the measures are excessively limited or scant resources are mobilised. Both facts will inevitably lead to a failure to achieve full resolution of the existing problem, given the small scale of the measures.

Coverage

While the criterion of coverage is used above all in evaluating the implementation and the results , it may also be useful in the design stage, to detect the section of the population that benefits from the policy or programme, in comparison to all the people affected by the problem. A correct definition of the problem is therefore required in order to design a correct target population. At this stage it might be more related to the criterion of suitability (as well as relevance, possibly) than with implementation.

[11] *Understood as "the degree to which a programme reaches a part or all of the target population of a programme or policy" (AEVAL 2015).*

Internal coherence

The criterion of internal coherence assesses the relationship between the objectives of the intervention and the set of measures that have been designed and implemented. The measures and actions must be consistent with the set objectives within a logical, formal and rational structure. There must be a cause-effect relationship. If the objectives cannot be achieved with the designed and implemented measures, because they contradict each other, or because there is no scientific or real evidence that said measures will lead to certain results, then the programme or policy will lack internal coherence. At the same time, its internal coherence may present vertical levels (between objectives of a different level) or horizontal levels (between the objectives of different components of the programme).

External coherence

The criterion of external coherence refers to the actions that are performed in the area of other policies and whose aim is to solve certain aspects unforeseen by the intervention that is the objective of the evaluation, but is in line with certain facets of the same problem. External coherence is related to the multidimensionality of the problems and possible solutions.

Complementarity

Finally, complementarity refers to the need for the objectives of different interventions to be to coordinated or headed in the same direction, so that they produce the greatest potential benefits. And this will also prevent one of them from generating effects that annul, reduce or are contrary to the objectives of the intervention[12].

Evaluation questions

Evaluation questions are the basic unit of research, the "*system which will create the logical structure of the evaluation is based on a series of queries and hypotheses that will make it possible to implement the evaluation*" (AEVAL, 2010).

From a methodological point of view, evaluation questions are of great importance. The questions outline the scope of the evaluation and at the same time, lead to its design. They contain the focus and scope of the evaluation (AEVAL, 2015). The questions are the transfer or operationalisation of the evaluation criteria and in turn break down into evaluation questions and sub-questions. All questions that seek to research or respond to larger questions on the suitability of the intervention fall within this criterion.

[12] *For different institutions, external coherence and complementarity are synonymous, whereas for AEVAL the specific characteristics of public policies as opposed to programmes make it necessary to establish a distinction between the two evaluation criteria.*

Evaluation Matrix

Some of the essential methodological components of the evaluation are the criteria and the evaluation questions, that are integrated into the ***evaluation matrix***. The evaluation questions and sub-questions, the evaluation criteria (or elements for assessment that lead to a judgement), the indicators, the tools of analysis and the sources of verification are all included in the matrix. It is a means to ensure the thoroughness of the evaluation, its systematic structure and that it contains both the focus of the analysis and the dimensions to be analysed during the evaluation[13].

The following section shows an example of an evaluation matrix with the criteria for the design evaluation[14].

[13] *For greater detail regarding the components of the evaluation matrix, their characteristics and the methodological steps to be followed with regard to evaluation design and the matrices themselves, see AEVAL (2015).*
[14] *The evaluation questions that are listed in the matrix are not exhaustive.*

EVALUATION MATRIX

I. To what degree is the set of measures included in the intervention aimed at solving the main problems detected?

SUB-QUESTIONS	CRITERIA	INDICATORS	SOURCES	TOOLS
1.1 Have the requirements or obligations sought to be fulfilled by the intervention been clearly and explicitly identified?	**SUITABILITY**	Analysis of requirements and obligations Contextual analysis of the Plan	Documents, sector-based research, etc. Documentation regarding the intervention and the context	Studying the current problem Documentary analysis Interviews of relevant stakeholders
1.2 What are the causes that give rise to, condition, and/or modulate the problem or requirement?	**SUITABILITY**	Analysis of requirements, causes and origin of the problem	Documents, sector-based research or evaluations, theoretical framework of the problem, diverse bibliography, etc. regarding the need or problem	Variable, according to the target of the evaluation and the sector of intervention
1.3 What is the target population that suffers from the problem or displays the social requirement?	**SUITABILITY**	Analysis of the target population that suffers from the problem	Documents, sector-based research or evaluations, diverse bibliography, etc. regarding the need or problem	Variable, according to the target of the evaluation and the sector of intervention
1.4. What is the magnitude or intensity of the problem?	**SUITABILITY**	Analysis of the extension, magnitude and intensity of the problem	Documents, sector-based research or evaluations, diverse bibliography, etc. regarding the need or problem	Variable, according to the target of the evaluation and the sector of intervention
1.5 Is it a long-lasting problem and in what terms?	**SUITABILITY**	Analysis of the problem. Duration and characteristics	Documents, sector-based research or evaluations, diverse bibliography, etc. regarding the need or problem	Variable, according to the target of the evaluation and the sector of intervention
1.6 How does the context of the problem condition it?	**SUITABILITY**	Analysis of the problem with regard to the regulatory, economic, area-based, social, institutional and other contexts in which it occurs	Documents, sector-based research or evaluations, diverse bibliography, etc. regarding the need or problem	Variable, according to the target of the evaluation and the sector of intervention
1.6 What is the definition of the problem, the interests and strategies of the involved stakeholders?	**SUITABILITY**	Analysing the definition of the problem, the interests and strategies of the involved stakeholders Documentary analysis	Documents relating to the intervention design and management. Regulations Documents of the involved actors	Map of the stakeholders involved in the intervention and the requirements of different stakeholders Interviews of managers
1.5. To what degree do the changes to the current context modify or maintain the advisability of the intervention?	**SUITABILITY**	Contextual analysis	Monitoring reports Tables and checklists for monitoring	Documentary analysis. Semi-structured interviews Studies

II.	What was the importance of the measures taken within the general political framework?			
SUB-QUESTIONS	CRITERIA	INDICATORS	SOURCES	TOOLS
2.1 What is the relevance of each measure of the intervention in economic terms with regard to the rest of the implemented measures?	**RELEVANCE**	Analysing the indicators of main macroeconomic magnitude	Monitoring reports Official data provided by managers	Documentary analysis Semi-structured interviews of managers

III. To what extent are there other interventions that may resolve elements of the problem that have not been taken into consideration by this intervention?				
What is the degree of coherence and alignment between the objectives of different interventions?				
SUB-QUESTIONS	CRITERIA	INDICATORS	SOURCES	TOOLS
3.1 Are there other interventions that may resolve elements of the problem that have not been taken into consideration by this intervention?	**EXTERNAL COHERENCE**	Identification and analysis of other interventions or sector-based policies	Variable, according to the target of the evaluation and the sector of intervention	Variable, according to the target of the evaluation and the sector of intervention
3.2. Has the external coherence of the intervention with other ongoing plans or programmes been examined?	**EXTERNAL COHERENCE**	% of common measures between the intervention and other public interventions in any sphere of action. Analysis of the coherence of different interventions % of measures that overlap or are duplicated	Managers, monitoring reports, tables and checklists for monitoring Variable, according to the target of the evaluation and the sector of intervention Documents on implementation, guides,...	Documentary analysis Variable, according to the target of the evaluation and the sector of intervention Semi-semi-structured interviews of managers and involved stakeholders
3.3. Are the objectives of the intervention related to those of other public interventions, converging towards the self-same problem or requirement posed?	**EXTERNAL COHERENCE**	No. of measures of the intervention that are complementary with other public interventions No. of studies or analyses verifying the complementary nature of the measures of the intervention with regard to other public interventions	Managers Guides and documents	Documentary analysis Semi-semi-structured interviews of managers and involved stakeholders
3.4. Is there any inconsistency between the measures of the intervention and other interventions or National Strategies within the field of study?	**EXTERNAL COHERENCE**	Degree of coherence between the different interventions or National Strategies within the same field	Documents on implementation Variable, according to the target of the evaluation and the sector of intervention Statistical analysis	Questionnaire to stakeholders Variable, according to the target of the evaluation and the sector of intervention Coherence analysis
3.5 To what extent are the objectives in line with the objectives of other policies and instruments?	**COMPLEMENTARITY**	Identification and analysis of the objectives of different policies and instruments.	Monitoring reports	Documentary analysis Interviews
3.6. Are there overlapping spheres of administrative competence between the different administrative levels?	**COMPLEMENTARITY**	% of common measures between the intervention and other public interventions in any sphere of action. % of measures that overlap or are duplicated	Managers, monitoring reports, tables and checklists for monitoring Documents on implementation, guides,...	Documentary analysis Semi-semi-structured interviews of managers and involved stakeholders

IV. What is the degree of coherence of the measures of the intervention with regard to the objectives sought to be achieved?				
SUB-QUESTIONS	CRITERIA	INDICATORS	SOURCES	TOOLS
4.1. Has the structure of the objectives, activities and results of the intervention been correctly defined?	**INTERNAL COHERENCE**	Degree of coherence between the objectives and the sought effects and results	Monitoring reports	Documentary analysis
			Variable, according to the target of the evaluation and the sector of intervention	Variable, according to the target of the evaluation and the sector of intervention
			Variable, according to the target of the evaluation and the sector of intervention	Semi-structured interviews of managers
4.2. Is the result that each strategic objective seeks to achieve, clearly identified?	**INTERNAL COHERENCE**	Indicators or definition of cross-cutting effects (synergies) of the measures, Relevance of the contributions for achieving the objectives of the intervention	Documents on implementation	Questionnaire to stakeholders
4.3. Is there any contradiction between the objectives or actions?	**INTERNAL COHERENCE**	Degree of coherence between the objectives and the actions	Statistical analysis	Internal coherence analysis
4.4. Is there a deadline for achieving the objectives?	**INTERNAL COHERENCE**	Evidence of the existence of objectives and time-based goals in the programming documents of the intervention	Tables and checklists for monitoring	Documentary analysis
		Periodic monitoring is the most suitable	Documents on implementation	Semi-structured interviews of managers and persons or bodies in charge
4.5. The definition of the different objectives responds to logical cause-effect relationships?	**INTERNAL COHERENCE**	Evidence of the logical definition of the causal relationships of the objectives	Monitoring reports	Documentary analysis
		Evidence of logical alignment in the level-based breakdown (from the strategic to the operational)	Tables and checklists for monitoring	Semi-structured interviews of managers
			Documents on implementation Statistical analysis	Questionnaire to stakeholders Coherence analysis
4.6. The indicators are suitable and specific to monitoring the achievement of the proposed objectives?	**INTERNAL COHERENCE**	Identification of the information in reliable databases	Official information of the Administrations and the managers of the measures	Documentary analysis
		Quality of the information compiled	Key stakeholders	Semi-structured interviews of key stakeholders

ANALYSIS TECHNIQUES IN DESIGN EVALUATION

For the design evaluation, there are different tools and techniques that allow the evaluator to obtain rigorous proof that responds to the evaluation questions or to analyse the different questions mentioned in the evaluation.

This Guidelines provides a brief description of the most relevant social research techniques that are of the greatest use and validity in evaluation. The most traditional classification of available techniques is that which distinguishes between qualitative and quantitative techniques.

Thus, among ***qualitative*** techniques, we have interviews, discussion groups, nominal group techniques, discourse analysis, SWOT analysis and case studies. And among ***quantitative*** techniques we have purely descriptive statistical analyses, of statistical inference or relations between the variables or phenomenon under study, either by means of statistical association or more complex analyses, such as simple linear regression models, multiple linear regression models, logistic regressions, etc.

When analysing the requirements or the existing problem, qualitative methods allows is us to obtain in-depth information on the perceptions and opinions of a groups of persons on a certain question.

These methods are normally supplemented with quantitative methods as they arise from the questions the design evaluation seeks to answer, and which are quantitative in nature, such as the quantification of the target population, or the intensity of the problem. The latter is mainly used in the evaluation of the results.

Figure 19. Analysis techniques in a design evaluation. Source: Author's own.

Type of technique		Purpose/nature
Qualitative techniques	Interviews	Exploratory. Applicable at any stage.
	Group discussions	Collecting qualitative information. Facilitating comprehension, credibility and acceptance.
	Nominal Group Techniques (NGT)	Structured analysis of ideas and problems.
	Discourse analysis	Analysing all discourses and the contexts in which they are produced.
	SWOT	Reducing uncertainty and define strategies.
	Case studies	Analysis of results and impacts.
Quantitative techniques	Survey	Obtaining descriptive information or other type of information in order to apply other techniques.
	Linear regressions	Analysis of explanatory causes and estimating effects.
	Logistic or probabilistic regressions	Analysis of explanatory causes and estimating effects.
	Cost-benefit analysis	Knowledge of differentiated impacts. Efficiency analysis.
	Cost-effectiveness analysis	Effectiveness analysis based on a relevant criterion.
	ARIMA Models	Time-series analysis.
	Multi-level analysis	Studying contextual factors, either by hierarchy or by levels.

Qualitative techniques

Interview

According to Dezin and Lincoln (2005), the interview is "*a conversation, it is the art of asking questions and listening to the answers*". This definition is based on a simple relationship between the researcher and the interviewee where the researcher asks questions that may range from opinion surveys or questionnaires, that is to say, highly structured instruments, to open interviews where the researcher may even been questioned or queried by the interviewee.

In qualitative research, the interview is not necessarily based on closed and structured questionnaires but on the contrary the researcher may repeat these meetings until all emerging or relevant topics have been clarified.

There are different types of interviews:

Structured interviews

In this type of interview, the questions to be asked are previously planned. A targeted and sequential list of questions is prepared. The interviewee cannot make comments or appraisals. These are closed questions; therefore the answers must be specific and exact.
Semi-structured interviews

The researcher prepares the questions beforehand on the basis of a thematic script. The questions shall be open and in contrast to structured interviews, the interviewee may express their opinions, qualify their responses, and even deviate from the initial script.

These are the most commonly used interviews in design evaluation.

Figure 20. Example of semi-structured questions in the Evaluation of the measures for streamlining and improving the management of Temporary Disability. Source: (AEVAL, 2009).

Evaluation questions	Evaluation criteria
What part of the evolution of the expenditure cannot be explained by the working population, the regulatory base, or by ageing?	Suitability
Have the General State Administration (AGE in Spanish) agencies been equipped with the organisational instruments?	Coherence
Is there complementarity and coordination of the intervention between the different entities responsible for managing temporary disability due to common contingencies and has the coordination been effective?	Complementarity and effectiveness

Unstructured or open interviews

These are generally known as in-depth interviews. In this case, the objective is to "to understand the interviewees' perspectives with regard to their lives, experiences or situations, expressed in their own words" (Taylor and Bogdan, 2008). These interviews are modelled after a conversation between peers and not a formal exchange of questions and answers. They require multiple meetings with the interviewees. There are there types of in-depth interviews: life histories, learning about events and activities that cannot be observed directly, and interviewing an extensive group. These three are of great use in applied social research but not directly in evaluation, as their goals are different.

Group discussion

Group discussion is a qualitative technique which brings together a group of people to obtain information on a specific topic, conducted by an interviewer.

Group discussion is a highly valuable technique to obtain information or qualitative evidence, as it generates a series of interactions among the people who are part of the group and it aids in obtaining information that is different from what is obtained in individual interviews.
When organising a group discussion, it is very important to be clear about the objective that is sought.

Based on each case, a group discussion may have different objectives:

- To share information and knowledge.
- To provide different perspectives.
- To find a common denominator.
- To come to an agreement.
- To compile qualitative information on perceptions, motivations, opinions, attitudes, etc.

There are different stages of development of a group discussion:

- *Establishing objectives*: The first step is to set the group objective and based on this decision, we shall define the type of group (more open or more closed) that we seek to form, whom to invite as participants (the sample), and develop tools for the group's functioning (script, schedule, activities, etc.).

- *Selecting participants*: In this stage we shall define the characteristics of all the participants and select the persons invited to form part of the group discussion. It is termed an "international" sample as it is not extracted on the basis of statistical criteria, nor is it a random selection, rather people are selected on the basis of their relationship with the topic under study.

- *Preparing the group discussion*: In this third stage, the group is planned, both with regard to the questions to be asked or the activities to be performed, and the logistic aspects.

- *Group organisation*: The group discussion is constituted.

- *Analysing information and drawing conclusions*: In this last stage, conclusions are obtained from the observations and results of the group work.

There are different types of group discussions, depending essentially on the role adopted by the group moderator and the level of conducting; from very open groups where different members of the group participate in a debate on the basis of pre-set questions, to other more focused ones that apply specific group dynamics techniques and lead the group participants towards a concrete point.

Generally, group discussions fall into two large classes:

Focus group

It is a group session, conducted by a moderator. It consists of a debate between different persons based on a list of questions that have been defined in advance and where the moderator suggests issues or asks questions and the group participants respond to them.
The goal of this technique is to obtain in-depth information on a specific topic by listening to a group of persons related to the topic under analysis.

Group dynamics

In this case, we are dealing with a programmed session with a series of activities and specific group dynamics that seek concrete objectives.

The objective of this technique is highly varied, although it focuses on analysing and diagnosing, or seeking symptoms and requirements of the analysed situation. Its goal is to propose alternatives and analyse the current situation with regard to certain envisaged objectives.

This technique has its advantages and disadvantages, as it helps to pool ideas, share experiences, and build consensus. It also helps to find the common denominator between the participants. Conversely, it may lead to organisation and logistics problems and it requires prior experience. Other disadvantages are that there may arise problems, arguments, and complaints that the moderator may not be able to control.

Nominal Group Technique

Nominal Group Technique (NGT) is a creative technique for analysing problems that combines individual opinions and facilitates the decision-making process. It helps to identify the elements of a situation or problem, gives partial or total solutions to them, and establishes priorities by consulting a group of persons while respecting their anonymity.

Its development consists of five stages:

- *Formulation stage*. In the first stage, the questions are posed linking them to the problems, obstacles, or difficulties.
- *Reflection stage*. In the second stage, all participants are asked to reflect on these questions silently and individually.
- *Grouping alternatives stage*. In the third stage, aided by the group participants, the researcher groups all the reflections made in the first and second stages, according to the degree of similarity of each, as judged by the group.
- *Debate stage*. In the fourth stage, a debate is initiated on the importance of each question that has been posed. The group votes on the groups of ideas.
- *Voting stage*. The fifth stage corresponds to the hierarchical arrangement of the alternatives.

The process concludes with the final report drafted by the expert, who passes it on to the relevant individual or body so that they adopt the required measures and attempt to solve the problems or questions posed in the NGT, or take into account the suggestions made by the participants.

This technique has a series of advantages, among them the systematic and orderly analysis of problems, as well as making proposals for decision-making by combining individual creative responses that become qualified group opinions.

When applied to public policies, this technique allows us to identify problems and their areas of improvement. It also lets us analyse their causes and solutions. In the Evaluation of the Human Resources Quality Plan of the General Services of the Administration of the Autonomous Regions of the Balearic Islands (ACAIB) (AEVAL, 2015), this technique was applied to three groups:

- The first group consisted of nine HR managers of the General Secretariats of the council offices of the Autonomous Region of the Balearic Islands (or CAIB in Spanish).
- The second group consisted of eleven heads of all the CAIB councils with a common denominator, they had staff members and at least four years of experience in public administration.
- The last group consisted of ten ACAIB civil servants. This group was characterised by its heterogeneity.

The methodology used was common to all three groups and it unfolded in the following manner:

1. Presenting the participants.
2. Formulating the first question. In your opinion, what are the main problems that affect the management of ACAIB personnel?
3. Silent generation of ideas.
4. Collecting the ideas-responses.
5. Group discussion of the ideas-responses, interpretation, and clarification.
6. Voting.
7. Break.
8. Formulating the second question: In your opinion, how can the management of ACAIB personnel be improved?
9. Session end.

Delphi Method

It is a group technique that allows us to classify expert opinions by means of an interactive process of individual questions.

It consists of four successive rounds of questionnaires. The responses are summarised in order to draft the next consultation and an agreement is reached.

After the first questionnaire round, we come to the next stage where the experts must again respond in view of the results of the first questionnaire and justify their differences with regard to the group. In the third stage, the expert is asked to comment on the arguments that deviate from the majority, and in the last stage, a final consensus is reached. The following section displays a brief overview of the process:

- *Early stage*: Defining objectives, identifying interviewees, and selecting the areas of study.

- *Development stage*: Designing and drafting the first questionnaire. Process and obtain the average of all the results. Identifying points of divergence and homogeneity. The results of the first questionnaire are used to draft the second questionnaire and so on and so forth.

Figure 21. Delphi Questionnaires. Source: Methodological Guidelines on Auditing for Inspectors of the General State Administration Services (December 2009).

	QUESTIONNAIRES		
	1	2	3
CONTENT	1 or 2 open questions	Transmit, prioritise and comment in favour of / against	Transmit and revise priorities
ANALYSIS	Classify, summarise and quantify	Identify areas of agreement / disagreement and establish priorities	Establish final results

- *Final stage*: The results are analysed and the conclusions drafted.

The benefit of this technique lies mainly in the insistence generated by presenting the same questionnaire several times. That is to say, the results of the previous questionnaires help experts to progressively learn about the different points of view so they may continue to modify their opinion if the arguments presented appear to be more suitable than their own.

SWOT Analysis

A SWOT analysis is a simple and general tool for taking strategic decisions. The main goal is to help find strategic elements and use them to make changes in the organisation by consolidating strengths, minimising weaknesses, taking advantage of opportunities, and eliminating or reducing threats.

It derives its name from the initials S (for Strengths), W (for Weaknesses), O (for Opportunities) and T (for Threats).

This technique is based mainly on two types of analysis, internal and external.

In an internal analysis, the objective is to detect the weaknesses and strengths of the organisation: to remedy the first and to promote the second. Different aspects are studied for this purpose: production, organisation, human or personnel resources and finances.

External analysis focuses mainly on detecting threats and opportunities. For this we shall consider the environment of the organisation, interest groups, legislative, demographic, and political issues. These points are very revealing when it comes to defining strategies that seek to combat threats and take advantage of opportunities.

Once the strengths, weaknesses, opportunities and threats have been identified, the SWOT Matrix may be created, which allows us to visualise and summarise the current situation of the organisation. With the results of the SWOT analysis, a strategy must be defined.
Below is an example of a SWOT matrix:

Figure 22. SWOT Matrix. Source: Author's own.

		STRENGTHS (S)	**WEAKNESSES (W)**
		Of the organisation	
OPPORTUNITIES (O)	**Of the environment**	Take advantage of the opportunities offered by the environment, using the organisation's strengths.	Take advantage of the opportunities offered by the environment, overcoming the organisation's weaknesses.
THREATS (T)		Use the organisation's strengths to avoid the threats posed by the environment.	By reducing the weaknesses of the organisation, we avoid threats.

The advantages of this technique are mainly that it leads to an awareness of existing problems, their characteristics and how they interact with the context, the organisation or the institutional framework, as well as the risks and opportunities generated by the environment that surrounds said organisation.

Quantitative techniques

Survey

A survey is one of the most frequently-used techniques in any type of evaluation, including design evaluation as it allows us to clearly identify the problems, the requirements, the magnitude or intensity of the problem, the perception of the stakeholders and the existing difficulties from the point of view of the stakeholders or the targets of the intervention. It also allows us to obtain results from a specific territory that may be generalised to the entire population. As a source of primary data, it allows the evaluator to arrange them in the most

convenient way possible to obtain the necessary information for the research.
It is a reliable but expensive technique and requires an exhaustive knowledge of the intervention and a thorough preparation of the framework of analysis by the evaluator.

When performing a survey, the first step is the ***sample selection***, which must be as representative as possible of the reference population, in order to make generalisations with regard to the population. Random sampling methods ensure the best sample representation. This means that any individual in the selected sample has the same probability of being selected.

Another aspect to be taken into account to optimise the results of the survey is the selection of the sample size. This requires a considerable knowledge of sampling techniques, a topic which is beyond the scope of this Guidelines. Nevertheless, it must be remembered that the greater the sample size, the lower the estimation error and thus, the more significant the results.

On other occasions, when the total population is not excessively high, all the members may be surveyed. Let us take, for example, a survey of organisations or units numbering between 100 and 200.

Once the sample size is selected, we come to the ***survey design***, which is the instrument for compiling and measuring data, and is characterised by a series of questions arranged according to a specific logic. Its design must be adjusted to the established objective and for this, we must be clear about what we wish to ask and above all, how to ask: It is important for the questions to be clear and concise, and flexible and "comfortable" answers must be provided to the interviewee. Finally, the questionnaire must not be very long.

There are different types of questions: ***open, closed, semi-open (or semi-closed)***.

With regard to the ***mode of administration*** of the questionnaires, they may be self-administered, in-person, telephone, postal or online surveys. The decision to opt for one or another depends on the advantages and disadvantages of each of them according to the topic under study, the available time and financial resources and the target population of the survey.

- In-person surveys are most frequently used in social research. They have the advantage of a more complete form of obtaining information and allow researchers to capture the environment surrounding the survey. But it has the disadvantage of being expensive, slow and difficult to access by certain populations.

- The main requirement for telephone surveys is that the surveyor must have a comfortable format. When drafting the questionnaire it is important to assess whether the design, duration, order, and interpretation are the most suitable. Currently they are mostly performed as Computer Assisted Telephone Interviewing (CATI), which lowers costs and the time required to perform them. However, it is not appropriate for delicate topics or complex questions. This survey mode may suffer from technical errors.

- In a postal or online survey, the interviewee reads the questionnaire and notes down their responses. There is no interviewer and therefore, a letter of presentation is required. It is a cost-effective technique and requires few personnel to perform the survey. It gives anonymity and flexibility of time to the interviewee. Its disadvantages include low levels of response and errors in filling out the questionnaire.

With regard to ***specific types of surveys***, we may mention:

- Omnibus surveys that allow us to include various topics, research or evaluation goals in a single survey. It is cost-effective, as instead of multiple surveys, only one is performed, thus sharing the research costs, and formulating a reduced number of questions in the same questionnaire and targeting the same sample. This type of survey is generally meant for large populations to achieve a financially feasible study. The questionnaire follows the same criteria as the interview but distinguishes itself by being arranged into different sub-questionnaires or modules with regard to different topics or outputs.

- Panel survey is a quantitative marketing research technique that is performed periodically on the same representative sample of a specific population.

Descriptive statistics

Once the survey is performed and the data has been filtered, they are analysed by means of ***descriptive statistical techniques***.

- Absolute and relative frequencies (the number of times an event is repeated and what it represents at the level of the population, respectively).

- Measures of centralisation are used (mean, median and mode) to obtain an overview of the data.

- Measures of dispersion, that provide an idea of variation in the sample data. They are useful when assessing the reliability of measures of centralisation such as the mean. They have an inverse relationship, the higher the measure of dispersion, the lower the representativeness of the measure of centralisation. The most well-known are variance and range; the range measures the difference between the maximum and minimum value that the observations can reach; variance measures the distance between the data and the mean.

These descriptive statistical techniques are characterised by their study of random phenomena; therefore their results are not precise and are accompanied by a certain degree of uncertainty. To measure this degree of uncertainty, we use statistical inference techniques.

Statistical Inference

Statistical inference techniques give us the answers to questions such as: What variables influence the incident? How do the variables influence the incident? Is it possible to obtain a model that explains the incident and allows us to predict its behaviour? Some of these techniques are described below.

Linear regression

In didactic terms, regressions seek to explain a variable or phenomenon that is deemed independent or endogenous by means of a series of facts, phenomena or variables that are called regressors, covariates or explanatory factors. It is the latter that may explain to a certain degree a phenomenon, behaviour, or reality.

Regression allows us to adjust a point cloud to a function where the endogenous or independent variable is explained partially through regressors or dependent variables, at the same time that the contribution of each dependent variable to the aforementioned explanation is determined. The difference between the real values and the explanation of the endogenous variable by the regressors is what constitutes the error term or random term.

When the independent variable is continuous and the function that links the endogenous variable with the regressors is linear, it is called linear regression. Apart from this configuration element, the assumptions on which the adjustment is made are: non-correlated regressors, their variance is constant (homoscedasticity), the errors in the measurement of each are inter-related and add to the total error, and the expected value is equal to zero, that is to say, the errors of a similar magnitude and opposite signs are equiprobable.

Provided the target of the analysis permits it and there is sufficient high-quality data, this technique can provide useful evidence for an evaluation.

Example: Evaluation of the Plan for Measures to Improve Cross-Border Health Services (AEVAL, 2013).

The third step is to apply the personnel estimation model. For this, a linear regression model has been developed that estimates the staffing of each service that would correspond to its calculated complexity, and identifies the services that exceed or fall short of said estimate. The dependent variable considered when building the model is the total occupied personnel on 31 December 2012, and as sub-group, the inspectors (both A1 and A2). The independent variables or predictors are the total complexity of the services and the total number of entries (records) in groups of a thousand. Additionally, dummy variables are created for the qualitative variables of time and service so that they are considered when calculating the estimate. Of the models built, the one with the best statistical adjustment has been selected.

Logistic or ordinal regression

Linear regression is a regression module where the variable or fact to be explained takes either two values (the phenomenon takes place or it doesn't, i.e., yes or no) or very few values (for example a scale of 5 values that measures intensity as a lot, enough, little, or nothing). Or to put it in another way, the variable to be explained is not continuous or the function is logistic. Similar to linear regression, logistic regression allows us to adjust a cloud of points to a function where an endogenous variable is partially explained through regressors.

Cost-benefit and cost-effectiveness analyses

Before assigning monetary resources to a public or private intervention, the quotient of discounted cash flows between the allocation of resources (cost) and their returns (profits) allows us to assess in absolute terms the convenience of allocating said resources or eventually of allocating them to alternative options. Occasionally, when the costs of the evaluated event are not explicit owing to the fact there is no market that reveals them, the so-called shadow prices are adopted as prices that they would have under perfectly competitive conditions.

The cost-effectiveness analysis is a variant of cost-benefit analysis that is applied when there is a lack of prices to assess the target or set of targets that the intervention seeks to achieve. To this end, cost would be that which allows the maximisation of the target. When alternative interventions to achieve the same target are compared, the selection criteria shall be to consider the intervention that helps to reach the target at a lower cost and at equal costs, helps to maximise the target.

Whenever faced with a problem that is resolved by cost-benefit or cost-effectiveness analysis, these techniques may constitute evaluation criteria.

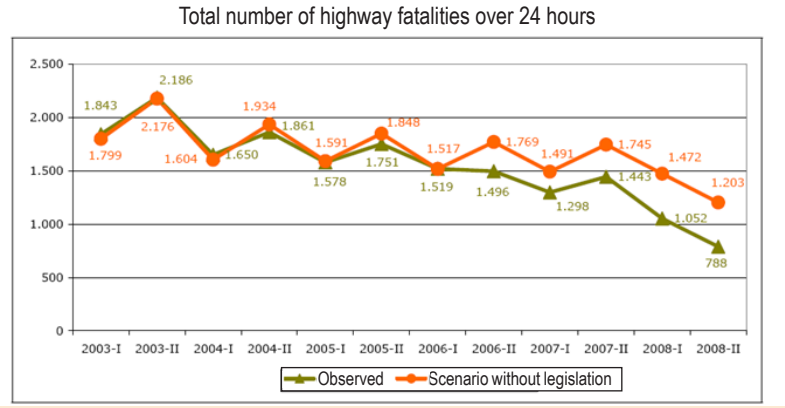
ARIMA Model

The ARIMA model (*Autoregressive Integrated Moving Average*) is a technique used to establish patterns of behaviour or of facts with the goal of forecasting. It does not use other variables as in regression techniques, but past data or values. Each observation is determined by earlier values in time. The ARIMA model (p.d, q) is denoted by means of three parameters - p, d, q, non-negative integers - that highlight the order of the three parts of the model: autoregression, integration and moving-average.

ARIMA models are used in evaluation to define patterns and make predictions. It is a dynamic time-series model, that is to say, future estimates are explained by the data of the past and not by independent variables.

Example: Evaluation of the Strategic Plan for Road Security 2005-2008 (AEVAL, 2009).

This evaluation uses an ARIMA model to study the impact of certain key variables on the victims' time series, primarily the legislative changes generated by the Plan (points-based driver's license and reforming the Criminal Code, above all). The study highlights that although there was already an underlying cause that implied a descent in the number of fatalities, what is certain is that "the impact of the plan and especially, of the plan put into motion from 2006 (especially the points-based license and the Criminal Code reform) has been responsible for reducing almost all fatalities over 24 hours". The following figure displays the differences between the observed situation (green line) and that which would have occurred without the implementation of the measures according to the ARIMA model (orange line).



Multi-level analysis

Multi-level models of analysis (hierarchical linear models, linear mixed-effect and nested models, among others) are models with parameters that vary in more than one dimension. They are of use when discerning what part of an effect may be attributed to one cause and what part to another, when both are present at the same time.

For example, in research on education, they would be useful to measure what part of the students' performance is due to the teaching method or to the school of institution where they study, and what part to other variables such as the social background of the students.

Frontier or efficiency models

Another tool that helps us to analyse certain phenomena in terms of efficiency or inefficiency of the resources used with regard to the maximum potential results that may be obtained with them. These are frontier analyses of the production or cost function. Based on the definition of a Production–Possibility Frontier (PPF), these models display, firstly, the parameters that define the frontier by their functional characterisation and subsequently, the efficient options

(the ones that are situated on the production frontier) and the inefficient ones. An allocation of economic resources is efficient when it is situated on the PPF.

There are three types of frontier models, of which essentially two are important:

- Nonparametric or mathematical models. This is data envelopment analysis (DEA). It uses mathematical programming to establish the set of observations that estimates the frontier and which do not require a previous functional form.

- Parametric or stochastic frontier models. It allows the estimation of the frontier functional form, costs, or benefits, the parameters, and its advantage is that it incorporates the specification error and allows us to distinguish the effects of noise or inefficiency error.

Stochastic frontier models are included in evaluation as an analytical option for applying the efficiency criteria.

Factorial analysis and principal component analysis (PCA)

When faced with a high number of variables with different degrees of correlation or linear dependency between them, both techniques may be used to reduce them to a set of factors or components that provide a synthesis of the phenomenon under study. Principal component analysis and factorial analysis both reduce the number of explicative variables, but differ in how they do it.

In the case of factorial analysis, the original variables are grouped by factors, so that they may be defined as linear combinations of the factors and explain the covariance or correlations between them.

Conversely, principal component analysis (PCA) defines new variables or linear independent components from the original variables. By means of a linear transformation, it defines a new system of coordinates for the original dataset where the highest variance is assigned to the first principal component, the second highest variance to the second component and so on, until the total variance contained in the original variables is saturated. In PCA, components are calculated as linear combinations of the original variables, normally after centring the data in the average of each.

Both techniques may be used in evaluation for exploratory, analytical, or confirmatory purposes.

Impact evaluation methods

Impact analysis or evaluation methods allow us to determine what part of the observed effects or results of a phenomenon are solely and exclusively attributable to a fact, in this case, a programme or an intervention. They are also called counterfactual methods.

The advantage of this type of methods is that they statistically isolate multicausality and isolate the effects, so that we may state with statistical rigour that the observed results are the result of a factor, fact, programme, or intervention.

Impact evaluation methods compare the results observed in the population, drawing a distinction between the target group or persons who receive an intervention and those who do not receive it, called the control group. If both groups are statistically similar or identical, the observed result can only be dependent on the treatment.

Impact evaluation tools or techniques may be divided into those based on experimental models, when it is possible to define in advance the phenomenon that receives or does not receive the intervention, through random processes; and quasi-experimental models where it is not possible to randomise in advance.

Multiple criteria analysis

Occasionally, the target of the evaluation may be assessed according to various criteria. On the basis of the weight of each criterion and according to a ratings scale, it is possible to quantitatively measure the joint application of different criteria and to sum them up in a number (the sum of the products: the weighting applied to the criterion by points attributed to the criterion), and thus compare alternatives.

LIST OF FIGURES

WORKS CITED

AEVAL. (2009). Evaluación de las medidas de racionalización y mejora de la gestión de la Incapacidad Temporal. Madrid.

AEVAL. (2009). Evaluación del Plan Estratégico de Seguridad Vial 2005-2008. Madrid.

AEVAL. (2009). Evaluación del Plan Nacional de Transición a la Televisión Digital Terrestre (TDT). Madrid.

AEVAL. (2010). Fundamentos de evaluación de políticas públicas. Madrid. Obtained from http://www.aeval.es/export/sites/aeval/comun/pdf/evaluaciones/Fundamentos_de_evaluacion.pdf

AEVAL. (2011). Evaluación del Plan Español de Energías Renovables. Madrid.

AEVAL. (2011). Evaluación del Programa de Ayudas a las Actuaciones de Reindustrialización (REINDUS). Madrid.

AEVAL. (2013). Evaluación del Plan de Medidas para la Mejora de los Servicios de Sanidad Exterior. Madrid.

AEVAL. (2014). Evaluación del Plan de Ahorro y eficiencia energética 2008-2012. Madrid.

AEVAL. (2015). Evaluación del Plan de Calidad de los Recursos Humanos de los servicios generales de la Administración de la Comunidad Autónoma de las Illes Balears. Madrid.

AEVAL. (2015). Guía práctica para el diseño y la realización de evaluaciones de políticas públicas - Enfoque AEVAL. Madrid.

AEVAL. (2017). Evaluación intermedia del Plan Nacional de Desarrollo Rural (PNDR). Madrid.
Aguilar, L. (1993). Problemas públicos y agenda de Gobierno. En L. Aguilar, El estudio de las políticas públicas. Antologías. México: Miguel Ángel Porrúa Editores.

Alvira Martín, F. (1991). Metodología de evaluación de programas. CIS.

Bardach, E. (1993). Problemas de la definición de problemas en el análisis de políticas públicas. En L. E. Aguilar, Problemas publicos y agenda de gobierno. México: M.A. Porrúa Editores.

Birkland, T. (1997). AfterDisaster: Agenda-Setting, Public Policy and Focusing Events. Georgetown University Press.

Bueno Suárez, C., & Osuna Bañeza, J. L. (2013). Reflexiones epistemológicas y metodológicas para la evaluación de políticas públicas. Andamios. Revista de investigación social.

CEPAL. (2005). Metodología del marco lógico para la planificación, el seguimiento y la evaluación de proyectos y programas.

Chen, H. (1990). Theory-driven evaluations. New Park, CA: Sage Publications Inc.

Chen, H. (1994). Theory-driven evaluations: Need, difficulties and options. Evaluation Practice, 15(1), 79-82.

Comisión Europea. EuropeAid - Oficina de Cooperación. (Marzo de 2001). Manual Gestión del ciclo del proyecto.

Denzin y Lincoln. (2005). The SAGE Handbook of Qualitative Research.

Dunn, W. (1981). Structuring policy problems. Public Policy Analysis Fih Edition. Pearson.
García Sánchez, I. (2010). Sistema de evaluación. Obtenido de http://www.eumed.net/libros/2010b(687/

Grau-Solés, M., Íñiguez-Rueda, L., & Subirats, J. (2011). ¿Cómo gobernar la complejidad? Invitación a una gobernanza urbana híbrida y relacional. Athenea Digital(11), 63-84.

Hoppe, R. (2018). Rules of thumb for problem-structuring policy design. Policy Design and Practice.

Instituto para la Evaluación de Políticas Públicas. (2019). Guía de evaluabilidad de intervenciones públicas. Madrid.

Ivalùa. (2009). Guía práctica 3. Evaluación del diseño.

Lima Facio, & Aguilar Astorga. (Abril de 2011). La importancia del diagnóstico en las políticas públicas. Contribuciones a las Ciencias Sociales. Obtenido de ww.eumed.net/rev/cccss/12
Linder, S., & Peters, B. (1984). From social theory to policy design. Policy formulationa and the challenge of conscious design.

Martinic, S. (1996). “Evaluación de Proyectos. Conceptos y Herramientas para el Aprendizaje”.
May, P. (1993). Claves para diseñar opciones de política. En: L.A. Aguilar (Ed.) Problemas Públicas y Agenda de Gobierno. . Mexico: MAPorrúa Editores.

Moore, M. (1993). Anatomía del problema de la heroína: un ejercicio de definición de problemas. En L. Aguilar, Problemas Públicos y Agenda de Gobierno.

Retolaza, Í. (2018). Teoría del cambio. Una brújula para orientarte en el camino. PNUD/Hivos.
Rossi, P. H., Lipsey, M. W., & Freeman, H. (2003). Evaluation: A Systematic Approach. SAGE Publications, Inc.

Ruiz Martínez, A. M. (2015). La participación ciudadana como criterio de valor en evaluaciones de carácter integral. Una aproximación desde la experiencia de AEVAL. GAPP(14).

Taylor, S. J., & Bogdan, R. (2008). La entrevista en profundidad. En C. d. docencia, Métodos cuantitativos aplicados. Antología. Chihuahua.

Vogel, I. (2012). Review of the use of theory of change in international Development. Review Report. Departamento de Desarrollo Internacional del Reino Unido.

Wildavsky, A. (1979). The art and craft of policy analysis. London: London and Basingstoke: Macmillan Press.