

Comparative study on the institutionalisation of evaluation in Europe and Latin America

Blanca Lázaro

Collection: **Studies n. 15**

Series: **State of the Art**

Area: **Public Finance**



Comparative study on the institutionalisation of evaluation in Europe and Latin America

Blanca Lázaro

Study n. 15

Series: State of the Art

Area: Public Finance



Published by:

EUROsociAL Programme
C/ Beatriz de Bobadilla, 18
28040 Madrid (Spain)
Tel.: +34 91 591 46 00
www.eurosocial-ii.eu

With the collaboration of:

International and Ibero-American Foundation for
Administration and Public Policies (FIIAPP).



This publication was prepared with the assistance of the European Union. The content is the exclusive responsibility of the authors and in no case should be considered a reflection of the views of the European Union.

Non-commercial edition.

Graphic production:

Cyan, Proyectos Editoriales S.A.

Madrid, July 2015.



Commercial use of the original work or any possible derived works, the distribution of which requires a license equivalent to that regulating the original work, is prohibited.

Table of contents

Prologue	5
I. Introduction	9
II. Conceptual framework of strategies for the promotion of evaluation.....	11
1. What is public policy evaluation?	11
2. What does institutionalisation of evaluation consist of?	15
III. Institutionalisation of evaluation in Europe	27
1. Evaluation in Nordic countries: the case of Sweden.	27
2. Evaluation in the Netherlands	33
3. Evaluation in the United Kingdom	41
4. Evaluation in Germany	47
5. Evaluation in France	52
6. Evaluation in Belgium	58
7. Evaluation in Southern European countries: Italy and Spain	62
IV. Institutionalisation of evaluation in Latin America.....	77
1. Evaluation in Chile	79
2. Evaluation in Mexico	87
3. Evaluation in Colombia	95
V. Conclusions and recommendations.....	107
Appendix I. Bibliography.....	117
Appendix 2. References consulted by country and source type.....	121

Prologue

For FIIAPP, a contribution such as this study to the existing social science literature on evaluation systems represents a significant advance in our founding mandate to cooperate in the strengthening of public systems by improving their structures, modernising their regulatory frameworks and policies, and training their civil servants and management officials. After starting out focussing on providing management capabilities to European Union twinning programmes, FIIAPP has gradually expanded its public assistance projects by mobilising experts and public administration officials to implement projects and programmes in over 90 countries around the globe.

During this time, FIIAPP has had the privilege of observing and playing an active role in the generation of public value in several regions of the world, which in turn has allowed us to compare and learn from our wide-ranging experiences. As things stand today, evaluation systems both in Latin America and Europe are immersed in a process of change associated with the regeneration of political systems, in all cases attempting to make public management more transparent and provide accountability to a public that makes use of new social participation channels. In some cases, specialised agencies have taken on the central role of articulating National Evaluation Agendas, whilst in other cases this task has been filtered through different ministries or offices responsible for budgeting, planning or spending, with institutions such as parliaments also playing an extremely diverse role in terms of the control and monitoring of public policies. In all of these cases, the systems are faced with the challenge of integration within new management models that enhance decision-making processes, incorporating elements that enable the development of policies, programmes and projects, and focus attention exclusively on the factors that make them pertinent, effective, and sustainable.

In fact one of the central pillars of the work of EUROsociAL II, the EU regional cooperation programme to promote social cohesion in Latin America, coordinated by FIIAPP, is to improve the quality of public spending through greater linkage between plans and budgets, and, above all, through the integration of evaluation in the public policy cycle. These are some of the main challenges facing the public administrations of Latin

America in orienting their management towards achieving results that satisfy the needs of their citizens.

The objective of the many different activities carried out has been twofold: the articulation and coordination of national, sectoral and regional plans with public budgets; and the incorporation of evaluation results in planning, budgeting and monitoring processes.

In pursuit of the latter of these two objectives, in July 2014 an International Seminar on Evaluation was held in Montevideo (Uruguay), where the participating countries exchanged experiences with the aim of together building a set of common basic guidelines for the institutionalisation of public policy evaluation. The result of the seminar, which was attended by representatives from Costa Rica, Ecuador, Paraguay, Peru and Uruguay, was the publication of the "Handbook on Good Practices in Systematisation of Public Policy Evaluation. Latin American Experiences".

As a way of providing continuity to this process, in March 2015 a workshop was held in Lima, Peru, involving representatives of key institutions in the management of public policy evaluation for these five Latin American countries. During the workshop, participants laid the foundations for sharing roadmaps and creating a Working Group to enable the mutual strengthening of institutions, both from specific areas and in general from other areas with similar responsibilities. The study presented here represents a contribution to the formalisation of this Inter-Agency Public Policy Evaluation Working Group, and serves as a support to the process of institutionalisation of the respective evaluation systems, both in Europe based on the experience of Latin America and in Latin America based on the European experience. We share a vision of formalised network models, producing sectoral evaluation methodologies and studies on strategic areas such as environment, infrastructures or public safety, providing the experience of the public administration and introducing new tools such as peer reviews of evaluations in order to generate new outlooks with which to face the changes occurring in a public action which is increasingly interconnected at regional level.

FIIAPP is committed to providing continued support for these tasks and in 2015, declared as the International Year of Evaluation by the Third International Conference on National Evaluation Capacities in São Paulo, has launched a series of actions aimed at promoting the management of evaluation as one of the central pillars of its work, introducing evaluation into the programmes that it manages as a necessary aspect for generating evidence and improving results. FIIAPP remains at the disposal of international bodies and national institutions that have a firm commitment to evaluation as part of the consolidation of the processes to which EUROsociAL

contributes today with this comparative study on the institutionalisation of evaluation in Europe and Latin America.

Pedro Flores Urbano

Director of the FIIAPP

I. Introduction

This report is intended to contribute to the description and analysis of the public policy evaluation systems being developed in various European and Latin American countries, and to the identification of factors that determine the proper implementation of these systems. Beyond the significance of the description and analysis themselves, the ultimate objective is to offer an approach to possible alternatives for the institutionalisation of evaluation in four countries of Latin America: Costa Rica, Ecuador, Peru and Paraguay.

The emergence, configuration and evolution of state-level evaluation systems has been the subject of analysis for a relatively short period of time, which undoubtedly represents a limitation for this study. This affects our understanding of what an evaluation system is; what its components are; what is its role and mission in a institutional public setting; its resources, activities, and the results to be expected from this system; the factors that determine whether it is functioning correctly or incorrectly, and whether its results are satisfactory or unsatisfactory.

Conversely, it is also important to observe the social and cultural setting in which evaluation is carried out. Successful recipes for a specific context may not work in a different one since they will need to be implemented by other people who may be immersed in a different political, administrative and organisational culture; under social and economic conditions that may vary substantially; and as part of personal, professional and institutional networks that may also be different.

This report has been prepared based on a review of academic literature and documents produced by international bodies and public sector agencies in the countries studied. It does not include field work, particularly interviews with leaders or other stakeholders of the studied evaluation systems, and it has therefore not been possible to confirm or update the documentary information in some cases. With regard to Spain, the documentary analysis is combined with the experience of the author in the management of a leading body for policy evaluation in that country between 2008 and 2013.

Even with these limitations, the report provides the benefit of bringing together in a single document a panoramic view of both the evolution and current situation of evaluation systems in a good part of Europe and a selection of Latin American countries, as well of analysing the main issues, obstacles and lessons learned with respect to the creation and development of these systems.

The text is structured in four sections. The first section reviews and sets out the current conceptual framework for evaluation systems and strategies for promoting them. The second section focusses on evaluation systems in European countries. The third section analyses the situation in certain Latin American countries. The final section sets out conclusions and recommendations intended to serve as a basis for devising, in the short term, specific roadmaps for the institutionalisation of evaluation in the aforementioned four countries.

II. Conceptual framework of strategies for the promotion of evaluation

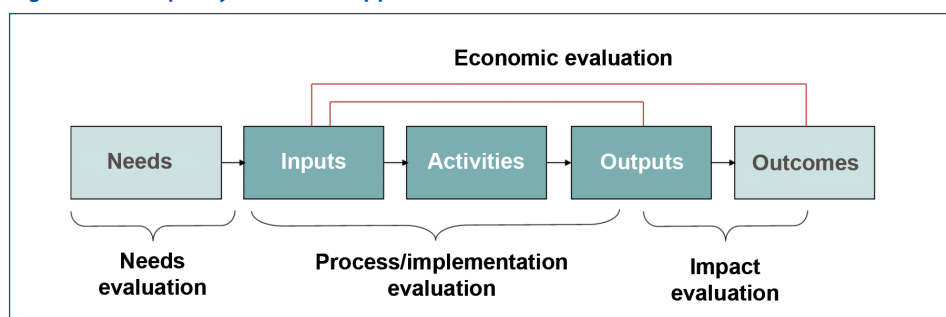
1. What is public policy evaluation?

Among the many definitions that exist in the literature on public policy evaluation, in this study we will use the definition included in the European Commission communication Responding to strategic needs: reinforcing the use of evaluation, according to which:

“...evaluation is a judgement of interventions according to their results, impacts and needs they aim to satisfy. It is a systematic tool which provides a rigorous evidence-base to inform decision-making.” (Commission of the European Communities, 2007, page 3)

This definition seems particularly appropriate because it includes different approaches to public policy evaluation (Figure 1) whilst also placing an emphasis on needs and impacts. In other words, it promotes as key evaluation criteria the extent to which the public intervention meets social needs.

Figure 1 Public policy evaluation approaches.



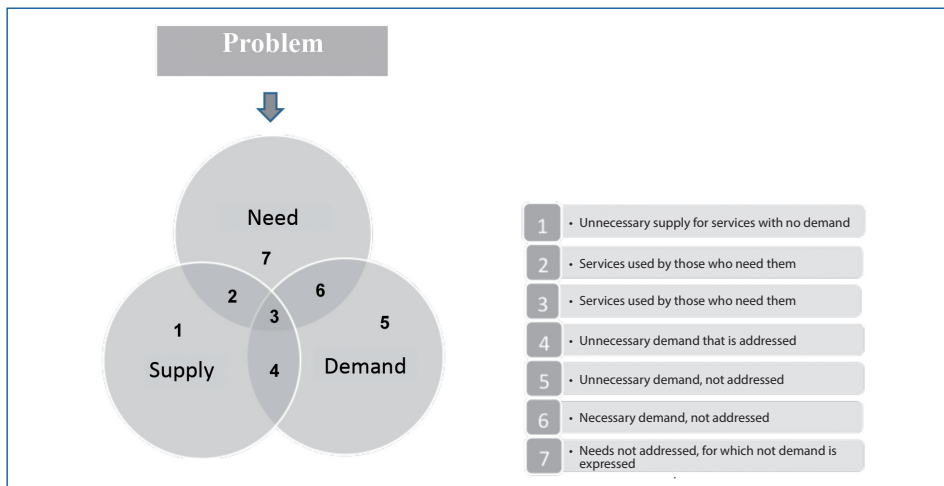
Source: Catalan Institute of Public Policy Evaluation (Ivàlua).

On one hand, this requires starting with an identification of the needs or problems (social, economic, etc.) that give rise to a public intervention. This is no mean feat,

given the conceptual and analytical challenge of defining and studying issues such as 'quality of life', 'social cohesion', 'competitiveness of the economy', 'public disengagement', etc. that are so often mentioned in political discourse as challenges to be addressed. Added to this difficulty is the confusion between needs and demand for public services, a confusion which often occurs in managerial approaches to administrative reform, mindful of the expectations and demands of 'customers' or 'users' of public services, but not necessarily meeting the needs of those who should be targeted (Figure 2).

On the other hand, according to this definition, public policy evaluation involves establishing the link or causal relationship between the public intervention and the intended changes with respect to the social needs, at both a theoretical level (based on an evaluation of the design of the intervention), and at an empirical level (based on the evaluation of effectiveness or impact) This causal analysis differentiates the evaluation from other activities that are often confused with it, such as the monitoring or follow up of public policies, audits and performance evaluation.

Figure 2 Need, supply and demand for services.



Source: Casado, 2009 (adaptation of William and Wright, 1998).

Let us take a brief look at the characteristics of this other type of activities in order to clearly distinguish between them.

- Performance monitoring has predominantly been carried out since the wave of public management and administrative reforms that occurred in the late 1980s under the paradigm known as New Public Management (NPM) (Osborne and Gaebler, 1992). Linked to management by results, these systems are designed to serve as tools to support political steering and the management and administration

of public sector organisations. They are normally used internally within public administrations, although they often include mechanisms for benchmarking against other similar organisations, ranking them based on certain performance or service quality indicators, etc.

One important difference between public policy evaluation and performance monitoring, other than the methodology, is that the former normally focuses on 'substantive policies', which are defined as

"...the 'normal' policies (such as social policy, employment policy, youth policy, housing policy) which essentially target the socio-economic situation and development in the policy environment" (Wollman, 2003, page 18).

Whilst performance monitoring focusses rather on aspects of implementation and outputs, often at public sector organisation level and not necessarily at public policy level. Evaluation and performance monitoring are not mutually exclusive, but complement each other and have obvious although under-exploited synergies. They are normally implemented and developed in parallel, as we will see. Table 1 summarises the differences between both activities.

Table 1. Differences between Evaluation and Performance Monitoring.

	Evaluation	Performance Monitoring
<i>Focus</i>	On public policies / programmes	On organisations
<i>Time</i>	Episodic, one-off evaluations	Regular, periodic practice, with monitoring at least once a year
<i>Methods</i>	Combination of quantitative and qualitative methods	Mainly quantitative methods
<i>Who promotes it</i>	Bodies or units responsible for public policy	Bodies or units responsible for budgetary control, organisation and/or HR
<i>Who conducts it</i>	External experts and/or internal evaluation units	Administrative bodies or units.

Source: created by the author based on Talbot (2010).

- **Public policy monitoring:** the differences between this technique and evaluation are similar to those mentioned for the case of Performance Monitoring, with the added nuance that, in this case, the monitoring function focusses on policies and programmes rather than organisations. Public policy monitoring therefore provides information on the status of the intervention and its context. This allows us to characterise the initial problem and how it evolves, as well as to determine how the public intervention is working, the extent of its coverage and its results, provided that the indicators are carefully selected, defined and measured. However, the monitoring does not include in-depth analysis of these different components, nor

does it allow us to determine the causal relationship between the intervention and the problem, as the effectiveness evaluation attempts to do.

- Performance Audit, Inspection and Oversight practices. Also originating from the NPM movement, these activities are aimed at controlling and verifying whether public sector organisations and their leaders have fulfilled the duties with which they are charged and whether they have done so through an efficient use of resources. These activities are therefore a type of analysis which is linked less to the effectiveness or impact of the activities or programmes than to the achievements of the organisations and their leaders, and to efficient use of the resources employed. These practices experienced strong growth during the NPM stage, as they were seen as a necessary counterbalance to the increased autonomy of public sector managers, which goes hand in hand with a greater need for accountability and control. These activities are usually conducted by National Audit Offices or equivalent bodies, which explains the increase in this type of institutions in OECD member countries, and particularly the rise in performance auditing activities from the 1990s onwards.
- Regulatory Impact Analysis is an ex ante analysis of alternatives and the possible effects of regulatory provisions in terms of their efficiency and cost-effectiveness (OECD, 2009). It pays particular attention to all types of costs that regulations can pose for the private sector, and particularly the business sector. The analysis techniques used include, among others, the Standard Cost Model.
- Accreditation of services and organisations. This activity is also very popular in OECD member countries, particularly in public service contexts such as higher education, prisons, hospitals etc. Accreditation is conducted by public sector bodies (e.g. university or research evaluation agencies) or private organisations (consultancies that implement and/or accredit quality management systems in various public services), within the framework of quality management systems.
- Finally, we should also mention mixed monitoring and evaluation (M&E) activities. These practices are designed to overcome the limitations of strict programme monitoring (lack of in-depth analysis, excessive focus on processes rather than results and impacts) and of ad hoc evaluations (specific studies not always adapted to a time frame and method of presenting results for maximising their use). The idea is to obtain a constant flow of information and rigorous knowledge in diverse forms—indicators, ex ante evaluations of implementation, impact, etc.—adapted to each phase of the cycle of the public interventions and their management and decision-making processes, including allocation of budget resources (Rist and Stame, 2011). Some countries in Latin America, such as Chile, are world leaders in development of these types of systems.

2. What does institutionalisation of evaluation consist of?

Institutionalisation of evaluation is understood as a process aimed at establishing an evaluation system in governmental settings through specific policies or strategies. In this section, we will first take a look at what we mean by evaluation system and then attempt to conceptually characterise the policies or strategies needed to promote this system.

However, we should first point out two limitations of evaluation, one technical and one political, to be taken into account in any strategy aimed at its successful development. Failure to do so may lead to false expectations about what evaluation can offer and about the role that it should play in a democratic political system, which could be counter-productive.

Limitations of evaluation

We begin with the assumption that the interest in promoting evaluation is based on its capacity to produce robust evidence on the operation, effectiveness and efficiency of public policies, which can then be fed back into decision making processes and thereby increase the effectiveness of public interventions.

However, from a technical-scientific standpoint, this production of evidence is a continuous process in which evaluation reports rarely provide fully conclusive evidence and should therefore not be considered the endpoint. On the contrary, each evaluation report typically responds to a series of questions whilst also posing others, suggesting programme designs or variants which, should they be put into practice, would need to be evaluated, or proposing enhancements to information systems which would allow future analyses to go further. In contrast, all too often evaluation reports are expected to give an emphatic and definitive opinion on the suitability of a programme. Such expectations can create frustration, which in turn can lead to a negative perception of evaluation and thereby hinder its progress.

On the other hand, from a political standpoint, the evidence produced by evaluation must unquestionably be made available to those making decisions with regard to the continuity, modification or elimination of programmes. However, in the words of Rist (2006), in a democratic system we should not aspire to occupy the place of honour (since this would be more in line with a technocratic-authoritarian system) but to be taken into account along with other legitimate criteria of a political nature.

In short, if we accept the characteristics and limitations of evaluation, the development and implementation strategy adopted must necessarily be accompanied by a process of change towards a different, more reflexive, administration. An administration that is more aware of the complexities of the social reality, as well as of its own limitations to

cope with them. An administration more inclined, therefore, to support the continuous production of quality knowledge, to promote well-informed political and social debate on possible intervention alternatives and their results, and to deal with failures and criticism as a form of learning. As Innerarity says:

'Democracy is a political system that creates deception... especially when it is done well. When democracy works well, it becomes a system of disclosure in which citizens oversee, discover, criticise, mistrust, protest and challenge.' (Innerarity, 2015).

Evaluation systems

The need to develop evaluation systems with a view to standardisation of the practice and its utilisation in governmental settings first emerged as a subject of research in the late 1990s through the works of authors such as Boyle, Lemaire and Rist (1999). These authors indicate that an 'evaluation regime' is one that is interested in the configuration of the capacity for evaluation, in the practice of evaluation, and in the associated organisational and institutional aspects.

Furthermore, according to Furubo and Sandahl, an evaluation system exists

"...when evaluations are no longer commissioned and conducted on an 'ad hoc' basis but through more permanent arrangements, which aim to guarantee, in advance, the supply of evaluative information... [and] at the same time, the evaluations conducted are put to suitable use." (Furubo and Sandahl, as cited by Leeuw and Furubo, 2008, page 159).

Varone and Jacob (2004) elaborate on this idea of evaluation systems, as do Barbier and Hawkins:

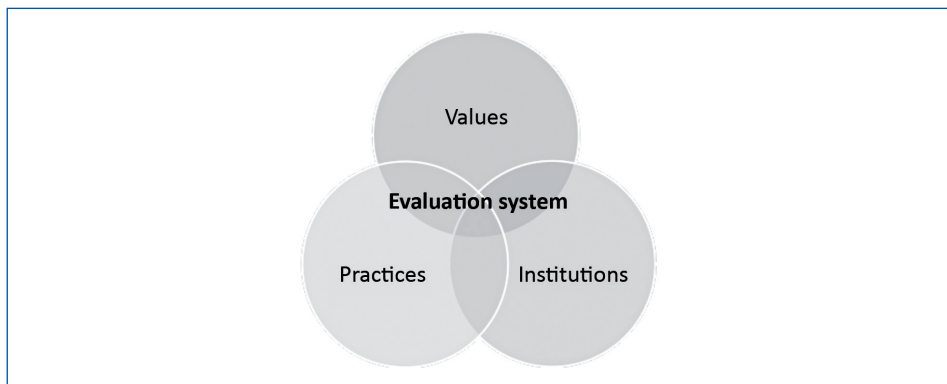
"...practices over time tend to be transformed into routines, especially when they succeed, and they are crystallized in formal institutions." (Barbier and Hawkins, 2012, pos. 158¹)

To summarise, we can define a public policy evaluation system as one in which evaluation is a regular part of the life cycle of public policies and programmes, it is conducted in a methodologically rigorous and systematic manner in which its results are used by political decision-makers and managers, and those results are also made available to the public.

Intertwined within a system such as this are the values, practices and institutions associated with a particular political and administrative system (Figure 3).

1. Note: as this is an e-book, we have replaced the page reference with a reference to the 'position' indicated on the lecture slide.

Figure 3. Evaluation systems.



Source: adaptation of Barbier and Hawkins (2012).

Furubo and Sandahl (2002) analysed evaluation systems in 21 countries and several international organisations based on a series of indicators that were subsequently revised by Leeuw and Furubo (2008) and which can be grouped together in the following categories: the existence of a common epistemological foundation; the existence of a culture that favours evaluation; the degree of institutionalisation, organisation and demand for evaluation; the supply of evaluation; the practice of evaluation and the degree to which it is used and disseminated for the purpose of improving programmes and accountability. Let us take a look at these distinctive features:

- The presence of evaluation in political, administrative and social discourse. Particularly, focused on the need for and utility of evaluation, on how to organise evaluation practices, how to use and disseminate evaluation results, etc.
 - The existence of a common epistemological framework: this refers to the need for consensus between the different agents involved in the evaluation system with regard to questions of what evaluation is and what type of knowledge it should produce (e.g. impact assessments or effectiveness evaluations; the use of experimental or quasi-experimental methods; performance auditing; project and programme monitoring; regulatory impact assessment; procedural compliance and/or attainment or objectives; the causal relationship between interventions and results, etc.; or how the evaluation should be conducted and for what purpose).
- In this regard, Jacob (2005) distinguishes different cognitive frameworks that encompass evaluation institutionalisation processes and which are normally found in combination, rather than in their 'pure state', in the different countries in which the process is observed:
- The classic 'control' approach of the administration.

- A 'managerialist' approach, based on performance monitoring and impact assessment (associated with the *New Public Management* movement, as mentioned earlier).
- A 'reflexive' approach, where evaluation appears as an alternative method of conveying information on public policies and public sector management, aimed at public sector decision-makers and agents involved in implementation of the policy, to beneficiaries and to the public as a whole.

This is a key point in the development of evaluation: the clarification of these frameworks and of the intention of each one, thereby allowing the creation of 'evaluation practice communities' that speak a similar language and share the same evaluation 'culture', as defined by Speer:

"... evaluation culture is ... a mainstream preference for specific evaluation practices, approaches and systems. Evaluation cultures develop and are influenced by the public policy sector and professionals and they are based on collective and operational choices as well as by rules-in-use." (Speer, 2012, pos. 1407)

- *Organisational responsibility*: there should be organisations (either one or several of them) that carry out evaluations, not just individual evaluators. However, there should not only be organisations that produce or are commissioned to produce evaluations, but also organisations that commission and use them, coordinate the production and use of evaluations, etc.

In this regard, Jacob (*op.cit.*) distinguishes between the various possible organisational configurations:

- *Monopoly*, where a single organisation has a dominant role in the promotion and practice of evaluation. The existing literature indicates that this system is associated with the initial stages of the institutionalisation of evaluation.
- *Centralised Pluralism*: within the context of an organisational mechanism for evaluation, a body and/or set of rules plays a predominant role in determining the behaviour of all others, who voluntarily submit to this system. This configuration is quite common, especially in mature systems in the Anglosphere².
- *Fragmented concurrence*, where the centralising element of the previous case is not present. This configuration may occur in countries that are strongly decentralised or in early or poorly developed stages of the institutionalisation of evaluation³.
- *Permanency*: evaluation activities should not be one-off, isolated activities, but should be conducted on a continuous and regular basis, with a certain volume and

2. Some important examples, such as the federal evaluation systems of the United States and Canada are outside the scope of this study.

3. From our point of view, this classification should include a 'decentralised' pluralist configuration, situated between centralised pluralism and fragmented concurrence, to cover developed evaluation systems that lack a central coordinating element. Whilst the fragmented concurrence scenario could be reserved for evaluation systems that are still weak and aspire, depending on the case, to a pluralist or centralised pluralist configuration but have not yet fully achieved it.

level of coverage. This also requires budgetary resources to be allocated in a specific and regular fashion.

- Emphasis on *utilisation*: evaluation results should be linked to decision-making processes and programme implementation processes (allocation of resources, adjustment of objectives or activities, etc.).
- *Public dissemination* of evaluation results, as a transparency and accountability measure. As Chelimsky points out:

"The government's responsibility to the people for its actions naturally includes telling them what it's been doing. And accountability also requires evaluators to report their findings to the public." (Chelimsky, 2008, page 402)

Moreover, public policy evaluation systems are not always articulated within, or necessarily in line with, the classic institutional frameworks associated with political systems:

- They can be developed around specific projects, such as transnational projects, as is the case with the Campbell Collaboration⁴ which centres on experimental or quasi-experimental evaluations and has led to the creation of its own evaluation system (Leeuw and Furubo, op. cit.), or projects within the field of international development, such as the International Initiative for Impact Evaluation (3IE)⁵.
- Alternatively, they can be developed around specific public policy sectors as in the experiences we will cover in greater detail when we review the situation of evaluation in several different European countries, namely: the Education Endowment Foundation in the United Kingdom; Fonds d'Expérimentation pour la Jeunesse in France; or, with a more disperse institutional architecture, the German evaluation system of active labor market policies launched after the Hartz reforms in 2003 (Jacobi and Kluge, 2006).
- Or, closely related to the experiences mentioned in the previous points, new trends that link innovation and impact evaluation, such as Social Impact Bonds, in the United Kingdom⁶ or United States⁷.
- In turn, several evaluation sub-systems may coexist within a single organisation. For example, this would be the case of the Office of Evaluation and Oversight (OVE) and the Office of Strategic Planning and Development Effectiveness (SPD) of the IDB; or the Independent Evaluation Group (IEG) and the World Bank's Development Impact Evaluation Initiative (DIME) of the World Bank, etc.
- Alternatively, within a single political system, several evaluation systems or sub-systems may coexist at state, regional or local level.

4. www.campbellcollaboration.org

5. <http://www.3ieimpact.org/>

6. <https://www.gov.uk/social-impact-bonds>

7. <http://www.whitehouse.gov/omb/factsheet/paying-for-success>

Factors that influence the development and evolution of evaluation systems

Furubo and Sandahl (2002) identify a series of determining factors in the development of evaluation systems:

- Firstly, *democratic quality*, and particularly the influence of transparency and participation practices which favour open social debate on public actions.
- Secondly, *scientific, technical and public sector management traditions*, which, if solid, will tend to favour rationalist approaches to decision-making. Otherwise, when the legal-administrative tradition is the dominant force in the administration and academic knowledge is scattered and dispersed, the progress of evaluation will be more difficult (Bemelmans-Videc et al., 1994).
In this regard, Jacob (*op. cit.*) adds the importance of favouring 'hybridisation' between administrative and academic cultures and practices, as well as between specialised areas of academic knowledge. Such hybridisation would be fundamental to the development of evaluation in public administrations.
- *The high level of public investment in certain sectors of the welfare state*, (or, in the current context, we could also add the peremptory need for adjustment) likewise favours the use of evaluation in decision-making.
- Finally, *the existence of driving forces exogenous to the system that are favourable to evaluation*. In particular, the influence of international donors and, in the case of the European Union, evaluation obligations associated with access to structural funds.
- Moreover, some authors also highlight the *institutional context and characteristics of the political system*. In this respect, according to Varone and Jacob (*op. cit.*), evaluation would be more easily institutionalised in democracies with majority rule electoral systems or presidential systems than in democracies with proportional representation or consociational systems, since in these latter cases the need to resort to strictly political criteria in order to reach agreements can diminish the role of evaluation in decision-making processes⁸. Under this same institutional approach, although this time in relation to *oversight bodies*, the same authors estimate that the degree of independence of these bodies with respect to the executive branch may facilitate institutionalisation. Furthermore, if they develop evaluation practices, this can also help to activate the oversight function of the parliament, which will either use the mentioned independent oversight bodies for this purpose or develop its own evaluation mechanisms.
- *Administration reform processes*, although, as we have seen in previous sections, in some cases such processes can promote a range of initiatives and do not always contribute to the development of evaluation. Wollmann (*op. cit.*) considers that the emergence of public sector reform processes is probably an important explanatory factor for the appearance of internal evaluation processes. However, with respect to

8. However, Switzerland, a country that does not fall within the scope of this report and is a model of consociational democracy, has a robust evaluation system.

independent external evaluations, Wollmann considers that the existence of institutional agents such as a parliament or court of auditors, and the maturity of the country's political culture, probably has a greater impact.

In order to truly appreciate the influence of some of these factors on the emergence of evaluation practices and systems in an international setting, we have provided an overview of the evolution of evaluation in Tables 1, 2, 3 and 4.

Table 1. Main stages of the expansion of evaluation. First wave: the expansion of the welfare state, interwar period and 1950s and 60s

Characteristics of the social and political context:

- Development of the welfare state in Europe and the United States. Growing complexity of governmental action and of public sector administration and policy due to the introduction of major public health, education and social assistance programmes.
- Centralised public sector management environment, with a strong presence of strategic and operational planning associated with the deployment of the welfare state and major anti-poverty, education, health, urban planning programmes, etc.
- Need for political legitimacy of public interventions in these new areas.
- Development of sciences and particularly social sciences and the analysis of public policies.
- Ideal of making the government more rational, more scientific and, ultimately, more effective at dealing with social issues.
- Separation of decision-making, policy conceptualisation and implementation. Evaluation affects the first, whilst the second is conceived in a mechanistic manner and has no implications to be taken into account for the evaluation and conceptualisation of policies.

Characteristics of evaluation at that stage:

- Focussed on analysing the results and impacts of substantive policies on education, social services and health, etc.
- Institutionalisation is (practically) non-existent.
- Practice of one-off studies commissioned from external academic experts.
- Use of eminently scientific experimental methods of a quantitative nature.
- Objective of informing the public sector decision-making process with regard to the most suitable mechanisms for achieving the goals set at political-administrative level.

Source: created by the author based on Furubo and Sandahl (2002) and Vedung (2010).

Table 2. Main stages of the expansion of evaluation. Second wave: halt to the expansion of the welfare state and advance of cultural and scientific relativism.

Characteristics of the social and political context:

- Deep economic crisis caused by an increase in oil prices from the early 1970s and throughout the entire decade.
- The repercussions of the crisis on public finances begins to erode the welfare state.
- International political instability (Yom Kippur War, Islamic Revolution in Iran, the establishment of dictatorships in several Latin American countries), although with some positives such as the transition to democracy in Spain.
- Start of deconcentration and administrative decentralisation processes in some countries.
- Erosion of the positivist paradigm in favour of constructivist approaches that deny the existence of an objective truth.

Characteristics of evaluation in that stage:

- Redefinition of the role of evaluation, which starts to be oriented towards reducing public interventions and maximising efficiency in the use of public resources.
- At the same time, we see the emergence of new demands for participation from social groups (young people, women and ethnic minorities) and public service users, and demands for further dialogue and communication in evaluations ("Dialogue-oriented wave" Vedung, 2010), amid vindications of improvements in the operation of representative democratic systems.

Source: created by the author based on Furubo and Sandahl (2002) and Vedung (2010).

Table 3. Main stages of the expansion of evaluation. Third wave: from the 1980s to the mid 2000s. Rise of neoliberalism and New Public Management

Characteristics of the social and political context:

- A chain of periods of economic boom and bust at global (74-75, 92, 2007), regional or sectoral levels (96-97, financial crisis of the "tiger economies" in South-east Asia; the dot-com bubble crisis, etc.), as well as at country level (economic restructuring in the UK, Spain, etc. in the 1980s) and a return to the welfare state.
- Successive enlargements of the European Union (+15, +25, +27).

Table 3. Main stages of the expansion of evaluation. Third wave: from the 1980s to the mid 2000s. Rise of neoliberalism and New Public Management (cont.)

- Increase in technical assistance programmes with a view to the incorporation of new countries into the European Union, as well as structural funds to promote development and equalisation of member countries.
- Wave of reforms in public administrations, known as New Public Management, amid questions with regard to the role of the state in society and the shape of public administrations and public services. Reforms have an unequal effect across OECD member countries (more intense in the Anglosphere) and also spread to other areas. In Europe it starts in the 1980s in the United Kingdom and Scandinavian countries. It later begins to spread to continental Europe in the 1990s and reaches Southern Europe by the 2000s.
- In very simple terms, under NPM, market forces and competition are generally the preferred means of providing public services more efficiently to beneficiaries, who are regarded as customers. The classic separation between political management and bureaucratic administration is diluted, with public sector managers being granted a high level of autonomy, which includes definition of objectives and the administrative instruments and management tools needed to fulfil them.
- Lack of confidence in the role of central planning. Deregulation, privatisation, efficiency and customer-orientation become the new buzzwords.

Characteristics of evaluation:

- Performance monitoring and measurement through the use of batteries of indicators primarily centred on the inputs, activities and outputs of organisations responsible for administrative functions and public services are used more generally as an instrument for ensuring accountability, whether between territorial agencies and central agencies, between central ministries and executive agencies, or between private providers of public services and the competent authorities.
- Evaluation of the effectiveness of public policies is a secondary concern. In contrast, an emphasis is placed on public service customer and user-satisfaction levels, benchmarking and the analysis of individual and organisational efficiency. Clearly the management aspect of NPM, which focusses on improving internal efficiency, tends to reduce the importance of evaluation.
- At the end of this stage, we see an increase in ex ante impact assessments and efficiency evaluations as a result of Better Regulation policies (Impact Assessment, Regulatory Impact Assessment).

Table 3. Main stages of the expansion of evaluation. Third wave: from the 1980s to the mid 2000s. Rise of neoliberalism and New Public Management (cont.)

- On the other hand, institutionalisation processes are started: in the European Union, mostly associated with structural funds and the European Social Fund, alongside administrative reform processes; and in Latin America, in some cases associated with the same processes, although also due to external factors, such as evaluation linked to large social programmes funded by international donors.
- In both cases, promotion and governance of these processes of institutionalisation is often provided by finance ministries or departments, which are responsible for creating and administering budgets, although in some cases also by presidential offices and/or sectoral ministries.

Source: created by the author based on Furubo and Sandahl (2002) and Vedung (2010).

Table 4. Main stages of the expansion of evaluation. Fourth wave: Late 2000s to date (2014)

Characteristics of the social and political environment:

- Deep global financial and economic crisis.
- State budget crises, especially in Southern European countries.
- Global wave of austerity policies and economic recession in various regions and countries.
- Heightened social inequality on a global scale.
- Institutional and social crisis, with increased questioning of the role of the state, the political class, mistrust of the private sector and in relation to collusion between economic interests and the political elite.
- Emergence of trends with critical views of NPM (New Public Governance, Public Values Governance, New Public Service, etc.) in the early 2000s which further intensifies towards the end of the decade. These new trends not only consider how to manage better but also the broader question of how to govern in increasingly diverse social environments, where there is a need to deal with increasingly complex issues that not only affect governments but also businesses, the third sector and society in general, amid an ever increasing level of public mistrust of public and private institutions.

**Table 4. Main stages of the expansion of evaluation. Fourth wave:
Late 2000s to date (2014) (cont.)**

- Within the OECD, the neo-Weberian (NW) model emerged with some strength, structured around certain public administration principles that guide the reform of administrations (Pollit and Bouchaert, 2004; SIGMA). We see a reaffirmation of the role of the state as the primary facilitator of solutions to the new problems posed by globalisation; the role of representative democracy as the legitimising element of the state apparatus; and the role of a revised and modernised administrative law in the preservation of principles that govern the relationship between the state and citizens.
- Other trends, such as Public Value Governance (PVG) (Bryson, Crosby, Bloomberg, 2014; also Chelimsky 2014), underline the predominance of democratic values beyond those associated with effectiveness and efficiency, and the special role of governments as guarantors of these values and as generators of public value, equivalent to the outcomes associated with the public interest. Public value emerges from inclusive dialogue and deliberation processes established among members of different sectors of the community⁹. Citizens move from being subjects of administration, voters, customers or beneficiaries of government action to actively participating in the resolution of issues of public interest and in the co-creation of public value.
- Moreover, this stage sees the promotion of the development of IT systems for public administration and statistics through the use of new technologies, as well as transparency policies and laws on access to public information and the reuse of administrative information.
- Linked to the above, there is an increase in the quantitative analysis of anonymised data from public records, accessible to all types of users (Big Data).

Characteristics of evaluation:

- Return to scientific-experimental models of evaluation, i.e. evidence-based policy making, although it is still far from being the dominant paradigm. This orientation has also started to spread to the evaluation of programmes financed by EU structural funds, particularly within the ambit of active employment and social policies (European Commission, 2014).
- At the same time, evaluation has spread to other fields of scientific knowledge, such as the environment, far beyond the traditional domains of social and health sciences.

9. "Public values and public value are not the exclusive province of government, nor is government the only set of institutions having public value obligations, [though clearly] government has a special role as guarantor of public values." (Beck Jørgensen and Bozeman, 2007, page 373-374).

**Table 4 Main stages of the expansion of evaluation. Fourth wave:
Late 2000s to date (2014) (cont.)**

- Rigorous impact assessments adopt “mixed method” approaches, applying qualitative analysis techniques in an attempt to better understand and explain processes and impacts.
- Certain level of institutional saturation in some cases (“inflation of evaluation”) with no clear performance levels.
- Emergence of the concept and practice of “social innovation” and “social investment”, linked to the practice of impact assessment.
- Collaboration with the private sector, which through “public-private partnerships” provides funds for experimental testing of innovative public or third-sector interventions, whether as philanthropic activities or investments.
- A wave of “flexible institutionalisation” of evaluation in Anglosphere countries outside the UE, particularly in the United States¹⁰.
- Emergence of “evaluation systems” and “evaluation policies” in international literature. Growing interest in research on institutionalisation trends and their results.

Source: Compiled by the author

10. For example, see the following memoranda on evaluation published by the Office of Management and Budget (OMB):

- Increased Emphasis on Program Evaluations (<http://1.usa.gov/bRGwej>), October 2009.
- Evaluating Programs for Efficacy and Cost-Efficiency (<http://1.usa.gov/aC2HSS>), July 2010.
- Use of Evidence and Evaluation in the 2014 Budget, (<http://1.usa.gov/KZ9fsR>), May 2012.

III. Institutionalisation of evaluation in Europe

1. Evaluation in Nordic countries: the case of Sweden.

1.1. General features of the evolution of evaluation in Nordic countries.

Evaluation in Nordic countries, especially Sweden and Denmark, has been developed in close connection with the establishment and expansion of the welfare state. The institutionalisation of evaluation has followed different trajectories and phases, and its degree of development varies from country to country.

Sweden is said to form part of the first wave of institutionalisation of evaluation, whilst Denmark, Finland and Norway form part of the second wave (Furubo and Sandahl, cited by Forss and Rebien, 2014). The evolution of the public sector and the professionalisation of evaluation have also followed different trajectories (Holm 2006), which has affected the demand, supply and utilisation of evaluations in different ways. Although one common feature is the fact that evaluation has a greater presence in the public sector than as a targeted academic discipline:

"...evaluation as a separate field of research and study hardly exists." (Holm, 2006, page 141)

Another common feature seems to be that, within the public sector, courts of auditors play a significant role not only in the auditing of public accounts but also in the field of evaluation, although not in terms of evaluation of the effectiveness or impact of public policies but rather in evaluating the performance of the executive branch:

"In all the Nordic countries auditing of state accounts is mentioned when discussing evaluation." (Holm, op. cit., page. 141)

In the executive branch, evaluations are generally conducted in a decentralised manner by the different sectoral ministries and evaluation practice is more deeply rooted in

certain areas of public intervention, such as in the case of educational, social, environmental and development aid policies.

Furthermore, the accession of Sweden, Finland and Denmark to the European Union also promoted the practice of evaluation, as in all other member countries.

As regards possible elements of an eventual common "culture" or "tradition" of evaluation, Forss and Rebien (*op. cit.*) present the following by way of an untested hypothesis:

- The fact that the role of the Public Defender or Ombudsman originates in Nordic countries indicates a strong tradition of evaluation within the public and parliamentary control system of the administration and the executive branch.
- A tradition of public hearings for draft legislation (something which is protected in the constitutions of Nordic countries) fosters a culture of involving stakeholders, linked to *ex ante* evaluation. This tradition of *ex ante* evaluation goes hand in hand with a strong tradition of *ex post* evaluation for amending and approving new legislation.
- The welfare system, financed with public funds (thanks to a system with one of the highest tax rates in the world) allows evaluation to play a leading role in ensuring a properly managed public sector.

In other respects, the Nordic countries coincide in the fact that they have small populations in quantitative terms; have been quite late in initiating modernisation processes; are located on the periphery of Europe; have large-scale welfare states and public sectors; and are multi-party democracies (Forss y Rebien, *op. cit.*).

Bohni, Nielsen and Winther (2014) compare the practice of evaluation in Nordic countries and identify a certain Swedish dominance, which could be due to the following factors:

- Firstly, the Swedish tradition of evaluation evolved more quickly and has a greater critical mass in terms of its scientific production and the quality of this.
- Secondly, the institutionalisation of evaluation in Sweden (and Finland) is, to a certain extent, different from the pattern followed in Norway and Denmark. More specifically, this is due to the existence in Sweden and Finland of internal evaluation units within government agencies, as well as the existence of national commissions with evaluation responsibilities. As such, these commissions and evaluation units in Sweden and Finland tend to commission evaluations from academic experts, whereas in Denmark and Norway there is a more pluralistic evaluation approach, with a greater level of participation of private consultancy firms, but also NGOs and other third-sector agencies.

1.2. Evaluation in Sweden

Evaluation in Sweden dates to the 1930s, although it became more robust in the 1950s, fundamentally centred on educational reform policies promoted at that time (Foss Hansen, cited by Holm, *op. cit.*). This development was determined by a political culture characterised by a long tradition (the oldest in Europe) of freedom of access to public information, as well as a political system that places great importance on dialogue and consensus (Kuhlmann, 2010).

According to Furubo and Sandahl (2002), two aspects especially stand out as determining factors for the development of evaluation in Sweden. The first is the deep-rooted practice of creating ad hoc commissions to analyse and decide on particular public policies. Evaluation has traditionally been an important input in the work of such bodies and in the decision-making processes they support. The second relates to the high degree of decentralisation of the Swedish government and to the conceptual and organisational differences between ministries (currently 13, employing 5,000 civil servants) and agencies (around 250, with some 200,000 employees) (ESV 2013).

This distinction reflects the principle of separation between the formulation of public policies, under the authority of the executive branch, and their administration and execution, the responsibility of agencies that have a high degree of autonomy with respect to ministries and therefore cannot exercise close control over their activities. In fact, agency budgets are independent of those of sectoral ministries and controlled directly by the Ministry of Finance. In contrast, agencies must be accountable for their activities and results to the Government, as a collegial body, which has facilitated broad penetration of the practice of evaluation and the utilisation of evaluation results.

In the 1980s and 90s, the development of evaluation activities in Sweden was affected by the successive economic and financial crises and by the reform processes in the public sector and administration taking place at the time. From the early 1980s, facing a important economic crisis that put its traditional welfare state at risk, Sweden embarked upon a twofold modernisation process. On one hand, in 1985 the Swedish government began a process of deepening its territorial decentralisation in a system that was already quite widely decentralised in comparison with other European countries. Furthermore, in 1988 and 1993 the country underwent a series of reforms in the area of public finance management, including the introduction of results-based budgeting, within the framework of the New Public Management wave. As such, the 'traditional' model of reform, aimed at a greater decentralisation of the government, with the consequent reduction of the powers and resources of the state's executive bodies in favour of local agencies, was combined with the introduction of typical NPM measures in the central government which especially affected the relationship between the central government and executive agencies, orienting these towards

results-based management. These same management systems were also introduced in municipalities, as we will see later.

On the other hand, in the early 1990s Sweden experienced a severe financial crisis, leaving it with one of the highest budget deficits in Europe (nearly 11% of GDP in 1993) and increasing public debt to 70% of GDP, (ESV, 2013). This crisis prompted a series of reforms within the institutional framework of management and control of public finances that have continued to present day and have, in turn, led to the implementation of sophisticated systems for analysis and monitoring of the management and use of budgetary resources.

Institutionalisation of evaluation in the central government

Evaluation practice is widespread and diversified in central government institutions in terms of the forms of provisions, to the extent that some authors refer to multiple "evaluation systems" that have been established over the years within public institutions in the areas of education, social policy, environmental policy and development aid, etc. (Holm, *op. cit.*).

On one hand we can see a system of commissions created on an *ad hoc* basis, as mentioned earlier, each articulated around a particular public policy. Although each case varies, these commissions usually include various different "stakeholders" ranging from representatives of political parties to trade unions and employer organisations, interest groups, associations, etc., depending on the context, as well as academic experts on the subject and public sector professionals. All these stakeholders have access to the same information to be used in the analysis and—probably more importantly and indicative of their degree of influence and involvement in the analysis tasks—participate in formulating the key questions that will guide the work of the commission and often refer to the efficacy and cost-effectiveness of policies.

Given the nature of their works and the type of material on which they are based, these commissions can be considered as the main transmission channel in Sweden between the existing academic knowledge on the different public policy areas and the corresponding decision-making process in public institutions. According to Furubo and Sandahl (*op. cit.*), although not all commissions include academic experts among their members, they often use academic materials as a basis for their work.

On the other hand, executive agencies have also taken on evaluation functions. In the first stage, evaluation activities were primarily conducted within the areas of education (by the Swedish Board of Education), energy and development cooperation (managed by the Swedish International Development Agency, SIDA), and were later extended to other areas of government.

However, there are no recent empirical studies that allow us to establish the level of coverage or the regularity of the practice evaluation among *ad hoc* commissions or executive agencies¹¹.

The evaluation activities developed by the executive branch have in turn been complemented by those conducted by independent bodies responsible for external evaluation of government policies, such as the Swedish National Audit Office¹², which publishes various Performance Auditing reports each year focussed on analysing the efficiency of government actions:

*"The starting point here is how well the subject of the audit is achieving its goals and the appropriateness of the organisation, operations, process, or function for the purpose. The goal of the audit is to make sure that the state, with regards to the interests of the general public, is getting good results for its efforts."*¹³

Numerous reports are published on this institution's website each year, many of which focus on analysing efficacy¹⁴.

Another important body is the Swedish Financial Management Authority (*Ekonomistyrningsverket* - ESV)¹⁵. The ESV is a central government agency attached to the Ministry of Finance and is responsible for efficient management of the budget. Its responsibilities also include providing agencies with tools for results-based management, auditing EU funds, etc., as well the ex post and prospective analysis of public finances.

Moreover, the Ministry of Finance also has other agencies that provide different information and analysis inputs on the use of public resources, such as the Swedish Office of Statistics (*Statistiska Centralbyrån*)¹⁶ and the National Institute for Economic Research (NIER)¹⁷.

In addition, parliament can also create specific commissions to evaluate certain public interventions.

11. On a visit to the publications section of the Swedish Government website(<http://www.government.se/sb/d/574>, on 1 February, 2015), only two documents containing the word 'evaluation' in the title could be found; neither document was recent and only one referred to a public policy: *Midterm evaluation of the Swedish Rural Development Programme 2007-2013*. Ministry of Enterprise and Innovation, Reports. 16 November 2010. The other was: *Evaluation of the Group Proceedings Act - Summary in English* Ministry of Justice, Information material. 28 October 2008. However, there may be additional documents available on the Swedish version of the site.

12. <http://www.riksrevisionen.se/en/Start/About-us/Our-fields-of-operation/>.

13. SNAO website, visited on 1 February, 2015.

14. Some examples for 2014: *Primary health care governance – based on need or demand?*; *Food control – is central government fulfilling its responsibility?*; *Utilising and developing the competencies of newly arrived immigrants – the right measure at the right time?*; *Effects of amended rules for the part-time unemployed*.

15. <http://www.esv.se/ESV/English/>

16. <http://www.scb.se/en/>

17. <http://www.konj.se/english>

In any case, as things stand today, evaluation is an integral component in areas such as budget, higher education and environmental policy.

Evaluation practices in local government

The penetration of evaluation practice in Swedish local governments has followed a "bottom-up" pattern, in which the Swedish Association of Local Authorities (*Svenska Kommunförbund*) has played a very important role (Kuhlman, *op. cit.*; Wollmann, *op. cit.*). The availability and quality of statistical information at municipal level in the Swedish Office of Statistics since the early 1980s has greatly facilitated the spread of these practices. This information includes numerous indicators and data on various locally-managed services and their costs (schools, social services for the elderly and people with disabilities, etc.).

The wealth of information available, the voluntary involvement of local bodies and the coordination of the association of municipalities has not only facilitated the analysis of each municipality but also comparisons between municipalities. Furthermore, over the years the indicators contemplated have been enriched and extended to cover the outcomes and impacts of municipal policies, as opposed to the initial phase when they primarily centred on more "conventional" aspects associated with costs and user-satisfaction (Wollman, *op. cit.*). According to some authors, public access to this information has a positive affect on its use by local authorities in decision-making processes:

"As performance indicators are accessible to all, they are used not only for the internal steering of the authority and the 'fiscal control of the businesses'. They are also taken seriously by politicians, particularly during elections." (Kuhlmann, *op. cit.*, page 343).

Professionalisation of evaluation

Sweden has a professional evaluation network that was set up in 1993 and subsequently re-established in 2003 under the name *Swedish Evaluation Association*. Much like other professional networks that exist in other countries and in the professional realm, the purpose of the Swedish association is to debate the role of evaluation and evaluators in society and in public institutions; to develop the professional competence of evaluators through guidelines, training and resources, etc.; and to establish links with similar organisations at the national and international level.

The evaluation system in Sweden

General assessment

Sweden has a consolidated evaluation system in both the central and local administration. That system has emerged thanks to Sweden's long-held tradition of transparency and

accountability, the great progress made in the social sciences since the interwar period, following in the footsteps of the United States, the quality of the Swedish democracy, and the development of its welfare state. In this context, and given the distinctive characteristics of the public institutional environment and particularly the configuration of the executive branch, the evaluation system is pluralistic and decentralised, although agencies such as the Swedish Financial Management Authority do coordinate the system to a certain extent. Likewise, agencies such as the Swedish Office of Statistics (*Statistiska Centralbyrån*) and the National Institute for Economic Research (NIER) provide know-how and information resources. Although we were unable to directly analyse a recent sample of evaluations during the preparation of this report, it is apparent from the documents and literature consulted that there are multiple approaches to evaluation, including rigorous impact assessments. Although public disclosure of evaluations is general practice, we have not had access to data that allows us to assess the degree of effective utilisation of results for the improvement of the programmes evaluated.

2. Evaluation in the Netherlands

In the case of the Netherlands, Frans L. Leeuw (2009) establishes that an explicit policy on evaluation in the central government began in the 1990s. However, evaluation activities had already made an appearance during the interwar period, again later in the 60s and 70s, and had gradually developed in the period from the end of that decade and throughout the 1980s.

It is rooted in the country's rich tradition of social research applied to public policies, and its progression from the late 1970s onwards is fundamentally explained by the budgetary pressures associated with the severe economic crisis in that decade, which intensified interest in establishing the impact of public sector programmes. Its emergence during the 1980s coincides with the introduction of New Public Management in the Dutch government and administration, together with a renewed desire for parliamentary oversight of the executive branch and the start of performance audits by the Netherlands Court of Audit (NCA), which revealed a lack of information on the effectiveness and costs of public interventions. Leeuw also points to the favourable administrative culture, which was exceptionally open to innovation and in which evaluation was welcomed as a tool for learning.

Origins

The origins of public policy evaluation in the Netherlands can be traced back to the 1920s when lines of research on social policy, and rural and urban development in particular, were being developed within the scientific community. Throughout the 1950s and 60s, certain researchers in this field took an interest in public policies, and

their research was largely commissioned and financed by the government, although they were in no way linked to budgetary/financial auditing, control or analysis, which was conducted by the different government departments as part of their normal duties.

The purpose of the research studies commissioned and aimed at public managers during this stage was to provide answers to questions about the causes of specific social problems and their possible solutions, but these were one-off studies, lacking a cross-cutting perspective and with a poor level of utilisation of results by those responsible for policy (Jacob, 2005).

In the 1970s, the pressures imposed on public finances by the international economic crisis drove the development of evaluation linked to the control of the budget deficit (Varone and Jacob, 2004). In the early 1970s the Ministry of Finance created an inter-departmental committee for the development of policy analysis (*Commissie voor de Ontwikkeling van Beleidsanalyse* – COBA) and charged it with promoting analysis and evaluation of public policy. That committee, which included representatives of all government departments, as well as of the Office of Planning and the Central Statistics Office, was assisted by a small team based in the Ministry of Finance, and one of its many activities was to publish a quarterly policy analysis journal (*Beleidsanalyse*) containing experiences from the Netherlands and other countries.

This initiative was carried out within the framework of the Planning, Programming and Budgeting System and Management by Objectives, and its objective was to provide an institutional basis for the use of public policy analysis and evaluation in the budget process. It was a question of ensuring that resources, instruments, objectives and the effects of policies were taken into account in management and the presentation of budgets.

In 1979, the committee published a report in which it highlighted political and administrative resistance to its progress towards its goals. More specifically it noted the difficulties derived from a multi-partite system which hindered the definition of clear objectives; the poor level of involvement of public officials in a process of reflection on its own work; as well as methodological issues. These difficulties seemed to spell the end of its work and of COBA itself in 1983.

However, this action line was continued within the Ministry of Finance with the creation of the Political Analysis Department (*Afdeling Beleidsanalyse*). Thus, despite the COBA ending in failure, it enabled the introduction of evaluation to the political agenda of the executive branch.

In the early 1980s, thoughts turned to the institutionalisation of evaluation. A report published by the Parliamentary Commission of Public Affairs (*Commissie Hoofdstructuur*

Rijksdienst) promotes experimental programmes and rules, and the practice of evaluation. The same Commission organised a survey of some 57 ministries, high-level public and parliamentary officials to assess the situation of evaluation. The results showed that few evaluations are conducted, that their quality was insufficient, and that the results are not adequately utilised. In contrast, over 80% of those interviewed believed that the institutionalisation of public policy evaluation was necessary (Jacob, *op. cit.*).

In 1982, one of the solutions recommended and implemented in order to control the public deficit was the "reconsideration procedure". This procedure focussed on the effects of policy but was ultimately intended to guide budget cuts. Be that as it may, it stimulated a strong demand for evaluation: 280 reconsideration studies were conducted by committees that benefited from the independence of judgement and action guaranteed by an agreement between members of the government stipulating that ministers would refrain from interfering in their work. From a comparative perspective, the example of the National Audit Office of the United Kingdom and the General Accounting Office of the United States was used as reference.

Moreover, the reconsideration procedure served as an opportunity to rebuild a bridge between the academic community and the administration. On this occasion, the involvement of high-level public officials was sought and not just the operational units, as was the case in the early years of evaluation.

Evaluation policy in the executive branch

The formalisation of policy evaluation per se began in 1991, after the approval of new legislation on governmental control earlier in the year (the Government Accounting Act), which established the obligation of ministries to evaluate their policies every five years. This formalisation also came after a brief document issued by the government (the Cabinet Memorandum on Policy Evaluation), which established the general principles governing evaluation practices. This document therefore laid to rest the idea of detailed regulations on evaluation which had been mentioned in the General Accounting Act and in other administrative provisions.

Among other things, the memorandum established the need for departments and bodies to create evaluation plans and to link the resulting information to their decision-making and budgets. The definition of methods, standards and other issues relating to the use of evaluation results was left in the hands of the different departments and agencies, although the creation of evaluation units to cover these functions was encouraged. Thus, this new regulatory framework allowed for the development and generalisation of evaluation practices in a large number of public bodies based on standards and procedures of homogeneous design.

Other milestones in the development of evaluation include the Government Act (1990) which stipulated a general principle of evaluation, and the Budgetary Procedure Act (1998) (*Van Beleidsbegroting tot beleidsverantwoording*—from budgetary policy to accountability) intended to link budgets to performance in public actions through the formulation of questions such as "What are the objectives? How are we going to achieve them? What are the costs involved? Have we achieved what we set out to do?" (Jacob, *op. cit.*). This procedure also imposed an obligation on each minister to publish an annual report describing the objectives and the achievements attained. The general intention was not so much one of sanctioning but of learning based on performance measurement.

It was a gradual process of accumulating experience in Dutch government ministries which, in 2002, led to the adoption of regulatory clauses of a general nature regarding evaluation (e.g. Regulation on performance analysis and evaluation, or *Regeling Prestatiegegevens en Evaluatieonderzoek*). These clauses paved the way for a more or less uniform and standardised application of evaluation practice across all ministries.

In addition, the package of reforms derived from rubric of New Public Management also saw the promotion of performance indicator systems throughout the 1990s in different areas of government, associated with the creation of departmental agencies that initially began in the 1980s and for which there was a renewed impetus in the 1990s.

Subsequently, in 2001 (the year in which the Dutch Evaluation Society was created) the Budget and Accounts Act was reformed, formally strengthening the role of evaluation. The new regulatory framework more clearly articulated the role of evaluation in the Dutch government, particularly in relation to the use of different evaluation approaches, the frequency with which evaluations are conducted, methodological issues (e.g. refining criteria for impact assessments), and other issues relating to the utilisation and dissemination of evaluations. Again, the executive departments and agencies were responsible for defining and developing these aspects in each case. To do so, they normally have internal inspection and finance units to conduct the evaluation activities (Wollmann, *op. cit.*).

According to Leeuw (*op. cit.*), general progress was noted, although there was a certain overlap of functions between public audit, inspection and evaluation bodies. Since the 2000s, this situation coupled with the administrative overload involved, as well as the public perception of persistent pockets of inefficiency, led the government to consider revising this current model, for which the results are as yet uncertain.

Generally, the institutionalisation of public policy evaluation in the executive branch and the Dutch administration is currently characterised by a search for practical solutions within the different departments, which benefit from the support of central

bodies that supervise the practice of evaluation. The Ministry of Finance stimulates and coordinates the mechanism, mainly through the creation of a network which stresses the link between evaluation and programme-based budgeting. The Ministry of Finance does not intend to take control of this mechanism, since it understands that evaluation is the responsibility of each individual ministry, although, just as the Netherlands Court of Audit, it can conduct meta-evaluations and other activities to supervise and check on how evaluation is being carried out. Its role is fundamentally a combination of providing encouragement, control and symbolic sanctions, calling for order where necessary (Jacob, *op. cit.*).

The role of parliament and the Netherlands Court of Audit

The involvement of the Dutch parliament began in the early 1980s based on conclusions of a study on industrial policy which were very critical of the government's activities. These conclusions clearly highlighted the fact that the enormous budgetary effort dedicated to the area had not provided results. That crisis opened a window of opportunity for parliament to create an internal evaluation office, however the opportunity was not seized. This can be explained, on the one hand, by the fact that relations between parliament and the executive branch in the country rested on the notion that the responsibility fell on the latter, which itself undertook to drive the evaluation of its activities; and on the other hand, by the complete independence of the Netherlands Court of Audit with respect to parliament and the executive branch (Jacob, *op. cit.*).

As a result of that situation and of the general confirmation of the lack of systematic monitoring of the effectiveness of public interventions, the NCA started to conduct government-wide audits on budget processes, policy implementation and use of grants. The purpose of these audits was to show the knowledge base available within government departments for acting effectively and efficiently. The conclusions were clear about the lack of knowledge available.

As regards the role of the Netherlands Court of Audit in the institutionalisation of evaluation in the Netherlands, in this country, as in Sweden, evaluation forms part of the core activities of this body, unlike in countries such as France and Belgium where these bodies have retained a more traditional role of controlling conformity and legality (Jacob, *op. cit.*). In the Netherlands, the NCA adopts a clear position on evaluation matters and is fully committed to supporting the actions conducted and establishing evaluation mechanisms. In fact, according to Jacob (*op. cit.*), it dedicates just as much of its efforts to evaluations as it does its traditional control functions.

Since 1991, following the adoption of the Dutch Government Accounting Act, the mission of the NCA was amended to place the emphasis placed on effectiveness and efficiency. This translates into the body conducting performance audits, defined as

checks aimed at evaluating how the entity under evaluation uses the public resources it receives to fulfil its mission within a framework of economy, efficiency and effectiveness principles.

Since the late 1980s, the Court has also been charged with the mission of controlling the quality of the evaluations produced by ministries, by conducting meta-evaluations. These evaluations are conducted for training purposes, to assist ministerial departments in developing their structures and evaluation practices. In addition, it is assumed that evaluation is carried out differently between the different government departments and, in the worst cases, the NCA is able to directly undertake performance audits seeking a pedagogical 'demonstration effect'. The existing literature on evaluation in the Netherlands contains references to studies of this type conducted by the NCA in 1990 and 1998 (Bemelmans-Videc, 2002; Jacob, *op. cit.*; Leeuw, *op. cit.*), which detected a great disparity both in terms of the number of evaluations carried out by the different ministries and in terms of their importance or the budget used. Identified as one of the possible causes of this situation was the lack of a clear definition of exactly what should be evaluated every five years. Following these studies, the Ministry of Finance undertook successive revisions of the evaluation policy, assisted by the NCA, which acted as a 'help-desk' for the different departments on this matter, whilst also monitoring the utilisation of evaluations. The level of utilisation seemed high, although the recommendations most followed were those affecting organisational rather than substantive aspects of policies (Jacob, *op. cit.*).

More recently, in 2012, the NCA prepared a new report on evaluation practice in the executive branch, referring to the period from 2006 to 2010, (NCA, 2012). According to this report, in the reference period some 1,200 public policy evaluations were conducted, 350 (29%) of which analysed the effectiveness of policies in terms of impact. The programmes evaluated represented a total budget of €51 billion, or 46% of budgetary expenditure in 2010. The study did not analyse the quality of evaluation, but it did identify errors in the classification of certain studies as impact assessments when in reality they were not, as well as a shortage of information to parliament on the evaluations conducted and, in particular, on the reasons for the failure to evaluate the effectiveness of certain policies.

In a subsequent report, the Netherlands Court of Audit detects hardly any improvement, confirming that the effectiveness of almost half of the policies funded by the central government had not been evaluated:

"There has been no change in the general picture that emerged from our initial audit. The effectiveness of slightly less than half of the policies funded by central government was not evaluated in the past six years... Opportunities are therefore being missed to apply the potential lessons learned in the formulation of new policy. It is important that the government has information on the effectiveness of policy in as many areas as possible. Public money will

then not be invested in policy that does not work effectively, if at all, and schemes that are effective will not be cut.” (NCA, 2013)

In order to resolve this situation, the NCA urged government departments to create evaluation programmes that reasonably establish what evaluations and what type should be addressed:

“We believe spending cuts and reforms should be implemented in a responsible manner based on reasoned choices. To this end, the government must have the relevant information it needs to determine the impact of spending cuts – not only budgetary information but also social information... The objective need not be to evaluate the effectiveness of all policy at all times, but every ministry should have a reasoned evaluation program laying down what kinds of evaluation should be carried out when and what kinds have the most added value.” (NCA, 2013)

At the time of writing this report, no subsequent reports have been published that allow us to determine the degree to which these recommendations have been followed.

As regards the development of evaluation at local level, some authors maintain that the Netherlands lags slightly behind countries such as Switzerland or France (Jacob, 2005), although development has been ongoing since 2002. However, the lack of empirical research in this regard at the time of writing this report prevents us from exploring this aspect in greater detail.

Professionalisation of evaluation

In the Netherlands, the evaluation community had long lacked a professional evaluation association (despite it being one of the founding countries of the European Evaluation Society), particularly after the boost provided by the management team of the NCA and the economic support provided the Dutch Ministry of Finance (Jacob, *op. cit.*). Subsequently, in 2002, a national association was created, the *Beroepsvereniging (VIDE)*¹⁸, which brought together inspectors, auditors and evaluators with the aim of promoting knowledge exchange at the national level to improve evaluation practice.

However, despite the relatively late creation of VIDE, evaluation started to be recognised as an academic and professional discipline in the Netherlands back in the 1990s (Bemelmans-Vidéc, 2002) and today is included in many university programmes at both undergraduate and graduate level. It also has a strong presence in scientific

18. www.videnet.nl

production and in the private consultancy sector, where there is a very active market for evaluation studies.

The evaluation system in the Netherlands

General assessment

The development of evaluation in the Netherlands is framed within a rich scientific tradition and an administrative culture that is very receptive to innovation and continuous improvement. After an early stage of one-off and external evaluations, a more systematic approach began to be applied following the economic crisis in the mid-1970s and the adoption of programme-based budgeting and management by objectives. These first steps were complemented by the strengthening of parliamentary oversight of the executive branch and the control duties of the Netherlands Audit Office in the 1980s. The evaluation system itself began to be developed in the 1990s through the entry into force of successive regulations establishing the obligation to evaluate, defining what evaluation should be understood to mean (including different approaches), attempting to organise evaluative practice through the creation of evaluation programmes, and establishing guidelines for linking the information outputs to decision-making and budgeting processes. The definition of methods, standards and other issues relating to the use of evaluation results was left in the hands of the different departments and agencies, although the creation of evaluation units to cover these functions was encouraged. This was a gradual process of accumulating experience carried out in parallel with the adoption of the monitoring and performance measurement techniques associated with NPM, whilst having no negative affect on evaluation. In this system, which we could characterise as a centralised pluralistic system, the Ministry of Finance played an important role in promotion and coordination, whilst the Netherlands Audit Office took on the role of general oversight of the system's operation. Its latest reports highlight some problems relating to the quality of the evaluations and show room for improvement in their use in decision-making processes.

3. Evaluation in the United Kingdom

The United Kingdom belongs to a cohort of countries that are pioneers in the development of public policy evaluation in the international context (Furubo and Sandahl, 2002). This can be attributed to the simultaneous existence of various factors identified in previous sections as determining factors for the development of evaluation. More specifically, the expansion of the welfare state after the Second World War, a long-held scientific tradition (including social sciences), and the quality of the political institutions and democratic culture.

However, the development of evaluation in the United Kingdom was brought to a halt by a wave of public-sector management reforms which began in the 1980s, in which evaluation was left with a merely residual role.

It wasn't until 1997 that evaluation returned to the political agenda, when it evolved in parallel with the performance measuring practices. There was almost no link between the two, and the role of evaluation remained a relatively subsidiary one:

"Evaluation has always been an underdeveloped part of the British policy system and despite definite increases in evaluative activity after 1997, under the rubric of "evidence-based policy," "performance" remains very much the predominant approach." (Talbot, 2010, page 1)

Performance measurement in the context of administrative reform

Over three decades, since the early 1980s, the different governments in power in the United Kingdom have promoted various policies for implementing performance measurement systems in the public sector. This process culminated in the late 1990s with the establishment of a practically universal system for measuring performance in the central government, local authorities, executive agencies and other public bodies, as well as non-governmental organisations providing public services.

This was possible thanks to the peculiarities of the British political system, characterised until very recently by a strong concentration of powers in the central government, with no constitutional recognition of powers for other levels of government. In terms of population coverage, after the processes of devolution to Scotland, Wales and Northern Ireland, the central government continued to control almost 80% of public service delivery (Talbot, *op. cit.*).

Important milestones in the reform process include the Financial Management Initiative approved in 1982; the Next Steps Initiative approved in 1988, which led to the creation of around 140 executive agencies; and the Compulsory Competitive Tendering (CCT) and Citizen Charters initiatives established in 1991.

The Financial Management Initiative of 1982 introduced a new philosophy of decentralised management, including budget management, in the central government which was gradually strengthened and complemented with other later initiatives. From the very beginning, this process involved the adoption of performance measurement systems, including setting objectives and indicators, as well as methods for measurement and regular reporting. More specifically, the performance measurement implementation and expansion process was characterised by three stages (Wollmann, 2003):

- Initial steps taken in the early 1980s in some local agencies and in the National Health Service (NHS), through ad hoc experiences.
- A more systematic development phase in central government between the late 1980s and the early 90s, during which certain tools and practices were imposed: Namely:
 - Key Performance Indicators (KPIs), which constitute the basis of the annual reports to be prepared and submitted by executive agencies to the government bodies to which they are attached, as of 1988.
 - Later, from the early 1990s, this same obligation was extended to other non-departmental public bodies (NDPB) and to the so-called *quangos* (quasi-autonomous non-governmental organisations).
 - From 1992 onwards, the central government established the obligation of local authorities to report annually on their performance, based on a system of 200 standardised indicators organised and supervised by the Audit Commission. Throughout that decade, similar performance measurement requirements were gradually extended to a wide range of public services such as the police, primary and secondary education, higher education, etc. In the mid-1990s there were hardly any organisations providing public services that were not subject to some kind of obligation to regularly report on their activities and results (Wollmann, 2003).

In 1997, New Labour moved for a strategic change to the system. In the central government, the use of Public Service Agreements (PSAs) was promoted and gradually extended in successive waves in 1998, 2000, 2002, 2004 and 2007, as different adjustments were introduced. The main difference between PSAs and the system introduced in the previous stage was the greater level of strategic orientation: PSAs expressed the medium-term public policy objectives of the different departments of the British government associated with multi-year budget allocations (three years) set through the use of Comprehensive Spending Reviews and Spending Reviews. Along these same lines, the measurement of progress was also meant to consider the outcomes or impacts achieved, although in reality this was not the case:

"In 1998 the aim was to cover all departmental objectives and to focus on 'outcomes'. Whilst the former was probably achieved, the latter certainly wasn't." (Talbot, op. cit., page 8).

In 2007, Gordon Brown's government introduced certain changes that brought a higher level of complexity to the system. On one hand, PSAs, which normally affected a single government department, became cross-cutting. At the same time, the number of PSAs was drastically reduced. Moreover, each PSA was complemented by Delivery Agreements setting out responsibilities and methods for the delivery of the public services in question. Alongside this, and as a result of the drastically reduced number of PSAs, each department was required to set Departmental Strategic Objectives (DSOs) for areas not covered by PSAs.

Evaluation

Evaluation practice in the British government was relatively low-level during the reform stage and up until 1997, in comparison with other OECD countries (Talbot, *op. cit.*). After 1997, the drive for evaluation by the Labour government translated into a range of initiatives, mostly promoted by special units attached to the Cabinet Office, the Office of the Prime Minister or HM Treasury (see Table 2). Traditionally situated outside the ordinary administrative structure, these units covered a range of functions, including identifying new ways of doing things and conveying that knowledge to the different public bodies.

The purpose of the Efficiency Unit, established in 1979, was primarily to improve the quality of public sector management in government, whereas the Prime Minister's Delivery Unit (PMDU), created in 2001, was focussed on improving the impact or outcome of public policies in different areas. The PMDU instigated various collaborations with government departments with the aim of setting objectives, learning to use data in order to understand and attempt to provide a solution to problems, producing regular reports, etc.

For its part, the Strategy Unit started in 2002, around the same time as the PMDU, was created by merging two other special units: the Forward Strategy Unit and the Performance and Innovations Unit. Its primary role was to identify and analyse cross-cutting issues that affected different government bodies and represented medium- and long-term challenges that required sophisticated analysis. It also assisted in the Spending Reviews initiative, and in budget planning, by helping to establish priorities. Within this framework, impact assessments were also conducted for specific projects in collaboration with sectoral ministries.

After the change of government in 2010, the Strategy Unit was dissolved and replaced by the Behavioural Insights Team¹⁹, which was initially created as a public body attached to the Cabinet Office but currently operates as a private entity jointly owned by the Cabinet Office and NESTA²⁰, and does not focus solely on public policies.

19. <http://www.behaviouralinsights.co.uk/>

20. NESTA was originally created by the British government in 1998 and is currently a private non-profit

For its part, the Government Social Research Service (GSRS), unlike the units mentioned above, is a kind of multi-discipline research and evaluation network with members in the main government bodies. This service has helped to promote evaluation and new evidence-based policy making trends. Among other deliverables, the service has published the Magenta Book²¹, a publication intended to serve as a guide for evaluators (on evaluation perspectives, methods, communication of results, etc.) which is periodically revised and updated.

Table 2. Special units attached to the government of the United Kingdom

Prime Minister	Cross-Cutting Units*	Thematic Units	Others
M. Thatcher (1979-1990)	Central Policy Review Staff Efficiency Unit		Next Steps Unit
John Mayor (1990-1997)	Efficiency Unit	Central Drugs Coordinating Unit	Citizen's Charter Unit
Tony Blair (1997-2007)	Performance and Innovation Unit (1998 – 2002) Prime Minister's Delivery Unit Strategy Unit (initially the Forward Strategy Unit)	Social Exclusion Unit (SEU), (transferred to the Office of the Deputy Prime Minister in 2001) Anti-Drugs Coordination Unit (transferred to the Home Office in 2002) Rough Sleepers Unit (Office of the Deputy Prime Minister) Respect Task Force (in the Home Office) Social Exclusion Task Force	Centre for Management and Policy Studies Office for Public Service Reform
Gordon Brown (2007-2010)	Strategy Unit Delivery Unit (transferred to HM Treasury)		
David Cameron (2010 to date)	Behavioural Insights Team (outsourced in 2013) Implementation Unit	Office for Civil Society Troubled Families Unit	

Source: Rutter and Harris (2014). * Does not include the Policy Unit created in 1974 and attached to the PM since then.

Under David Cameron's government, 2013 saw the creation of a network of What Works Centres²², which currently consists of nine centres specialising in education,

organisation. Its objective is to promote evidence-based innovation in various areas of public intervention (www.nesta.org.uk)

21. <https://www.gov.uk/government/publications/the-magenta-book>

22. <https://www.gov.uk/government/publications/what-works-evidence-centres-for-social-policy>

public health, prison services, local development and early social intervention policies²³. Taken together, the public policy areas covered by this network represent total spending of more than €270 billion. The objective of the network is to foster rigorous evidence-based decision-making. To this end, the centres produce evaluations and summaries of existing evaluations, and publish the results in a form that makes them easily accessible to public sector decision-makers, etc.

Finally, a recent report of the National Audit Office (2013) centred on economic evaluations and impact assessments highlighted unequal evaluation practice:

- Coverage was difficult to establish, given the lack of systematic information in this regard from government. However, the NAO states that between 2006 and 2012, government departments published a total of 6,000 evaluation documents on their websites, of which some 305 were impact assessments and 70 included cost-effectiveness analyses. Of these 70 evaluations, 41 affected a total of £12.3 billion of public spending (€16.5 billion).
- Variable quality: only 14 impact evaluations, from a sample of the 34 analysed by the NAO, had used robust methodologies to define the counterfactual and therefore presented reliable results. Furthermore, the methodologically weakest evaluations tended to show better results. In general, evaluations associated with education and employment policies showed a higher level of quality than those in other areas, such as urban policy or enterprise promotion policy.
- A lack of a connection between the results of previous evaluations and *ex ante* analyses or Impact Assessments associated with new phases of the same policies (only 40 out of 261 impact assessments published between 2009 and 2010 made any reference to evaluation results) and generally between evaluation and the decision-making process.
- The manner in which the evaluation function is organised varies from department to department and follows no clear pattern. Therefore, whilst internal evaluation units that either conduct or commission evaluations from experts are quite common, some departments have established bodies with varying degrees of independence to conduct all or part of their evaluations:
 - In 2011, the Department for International Development (DFID) established the Independent Commission for Aid Impact (ICAI), a non-departmental public consultation body reporting to parliament through the International Development Committee.

23. The National Institute for Health and Care Excellence (www.nice.org.uk); Early Intervention Foundation (<http://www.eif.org.uk>); What Works Centre for Local Economic Growth (<http://whatworksgrowth.org>); What Works Scotland (<http://whatworksscotland.ac.uk>); the Education Endowment Foundation mentioned above (<http://educationendowmentfoundation.org.uk>); the College of Policing (<http://www.college.police.uk>); What Works Centre for Wellbeing (<http://whatworkswellbeing.org>); the Centre for Ageing Better (<http://www.centreforageingbetter.com>), and the Public Policy Institute for Wales (<http://ppiw.org.uk>).

- The Department of Education created the *Education Endowment Foundation* (EEF)²⁴, which operates as a private foundation funded by a government grant and is responsible for evaluating educational interventions aimed at children at risk of social exclusion in England.
- Furthermore, according to the study published by the NAO, traditional barriers to quality evaluations continue to exist: problems of accessibility to administrative records; discrepancies between the pace and timing associated with decision-making processes and evaluations; a lack of demand for evaluations; etc.
- Generally, there is still considerable room for improvement in the use of evaluation results in decision-making processes.

The evaluation system in the United Kingdom

General assessment

The development of evaluation in the United Kingdom begins with 'assets' similar to those mentioned in the case of the Netherlands, namely the country's long democratic tradition, the strength of its institutions, its lengthy scientific tradition and, in particular, the developed state of social sciences. Although rigorous evaluations and the general interest shown by academia in studying public policies dates back to the interwar period, as it does in the United States, Scandinavia and the Netherlands, and remains strong throughout the postwar golden era, the advent of the oil crisis and the entry into power of neoliberal conservative governments mark the beginning of a protracted hiatus in evaluation practice in government, which did not end until the year 2000. Throughout these years, performance measurement practices associated with management by results and spending controls are imposed, reaching their peak in the late 1990s after having been extended to all central government agencies and local authorities. From 2000 onwards, public policy evaluation once again begins to gain momentum, driven by the Strategy Unit and the Government Social Research Service. Today, David Cameron's government continues to promote rigorous public policy evaluation through initiatives such as the network of What Works Centres. Moreover, there are various initiatives promoted by private non-profit entities, with government support, which are very active in the promotion of evidence-based policy making, such as the Education Endowment Foundation in education, or NESTA in innovation and development policies and also in local innovation. Nevertheless, despite the eruption of initiatives promoting evaluation, a recent report from the National Audit Office revealed the existence of areas for improvement associated with both the coverage and quality of evaluations, as well as the degree to which they are utilised.

24. <http://educationendowmentfoundation.org.uk/>

4. Evaluation in Germany

In Germany, we can distinguish three waves or stages in the development of evaluation (Speer, op. cit.; Derlien, 2002). The first was between the 1960s and 70s and was associated with the development of welfare policies (Hellstern y Wollmann, cited by Speer, 2012). The second wave began in the mid 1970s, in the midst of a significant economic and fiscal crisis, and therefore focussed on introducing criteria for guiding budget cuts and improving the efficiency of public spending. The third wave came in the early 1990s under the rubric of New Public Management, known in Germany as the New Steering Model, although in this country it was only weakly linked to evaluation.

From the mid 2000s to date, the country has been experiencing a fourth wave or stage in the development of evaluation (Speer, 2012). New laws have increasingly included clauses on evaluation and experimentation. These clauses are of limited duration, with the continuity of laws and agencies not guaranteed but dependant upon their performance. Such clauses have become widespread in the country's states or *Länder*. At the same time, evaluation systems and the flow of programme evaluations have been extending themselves to that level, particularly with respect to impact assessments.

Institutions

Germany is a federal republic comprised on 16 states or *Länder*. Balances of power and a distribution responsibilities exist between the federal and state levels which vary depending on the public policy area, which also affects how evaluation is carried out. This is the case with employment policies, which are designed at national level, and education policies, which are managed in a decentralised manner with the involvement of both the *Länder* and local governments, and where a great degree of diversity can therefore be observed, particularly with respect to evaluation practice.

Moreover, ministries enjoy a high degree of autonomy, which also influences the lack of a clear model for the institutionalisation of evaluation.

"Co-operation and consensus building are also key features of the way in which the federal executive works. The principle of ministerial autonomy means that the chancellery acts more as a co-ordinator than a driver of policy and law making. Centrifugal forces need to be kept in check and the system raises a significant challenge for the centralisation of reform, the establishment of clear reform leadership in the executive centre of the federal government, and the development of a collective, whole-of-government approach to reform." (OECD, 2010, page 60).

In Germany, evaluation has been implemented in a manner that is more procedural than institutional. National influence is primarily felt through laws and in the context

of the discretion of the different ministries. There is no over-arching institutional framework for evaluation at the state level. The Court of Audit (*Bundesrechnungshof*) has a limited effect on evaluation, since these activities are primarily conducted within the framework of sectoral policies, where different reforms and policy changes have led to a higher demand for evaluations and the funds needed to finance them. For its part, the parliament increasingly requires information on public policy evaluations, although interpretations of this information currently centre more on procedural issues (which evaluations are commissioned, which entities are evaluated, the evaluation schedule, etc.) rather than on the results of the evaluations themselves (Jacob, Speer and Furubo, 2015).

Despite the lack of a formal system for institutionalising evaluation, the existence of a mature evaluation culture, created and shared by a highly developed evaluation knowledge and practice community is undeniable. Indicators of this include the existence of DeGEval, a highly active German and Austrian evaluation association²⁵. We can also highlight *Zeitschrift für Evaluation*, a weekly evaluation journal, as well as the existence of evaluation standards approved by DeGEval back in 2001.

Practices

Public policy evaluations are conducted regularly in different public intervention areas. More specifically, in the areas of education and in employment policies, as well as in international development cooperation (Speer, *op. cit.*). That said, there is a far greater separation between the academic research sector and the administration in the evaluation practice in Germany than in other European countries. For example, employment policy evaluation is conducted by research institutes, although they often are attached to the administration.

Such is the case of the Institute for Employment Research (IAB)²⁶, a research institute attached to the Federal Employment Agency (*Bundesagentur für Arbeit* – BA). It was created in 1967 and employment policy evaluation is among its primary objectives. The objectives of the evaluations conducted by the BA through the IAB, include: improving the employment outlook and employability of the beneficiaries; evaluating programme costs and benefits; measuring the net macro-economic effects of programmes; and measuring the regional impact of employment policies.

Although under the BA, the IAB is an independent body and is free to publish evaluations as and when it sees fit. Its status as a research institute preserves its autonomy to do so and to apply rigorous evaluation standards and methods. The Federal Ministry for Employment did have its own evaluation unit, but this was

25. <http://www.degeval.de/home/>

26. <http://www.iab.de/en/iab-aktuell.aspx>

dissolved (Speer, *op. cit.*). Although the IAB conducts most employment policy evaluations, there are other agents, such as public research institutes and academic research organisations or groups, that also conduct evaluations in this sector.

As regards education policies, there are many institutions involved in evaluations in this policy area. For example, at the federal level, the Leibniz Institute for Educational Research and Educational Information (*Deutsches Institut für Internationale Pädagogische Forschung* - DIPF) is the leading organisation. At the *Länder* level, various ministries provide funding to research institutes that are often attached to local universities. Moreover, there have recently been cases of educational inspection agencies being created that are part of the *Länder* ministries. We have also seen the creation of the Institute for Educational Quality Improvement, a research institute affiliated with the University of Humboldt in Berlin, which focusses on improving educational standards.

The Ministry of Education has increased funding for educational research and for the evaluation of education policies. Schools have a budget for evaluation activities which, although somewhat limited in quantity, can be used flexibly for the evaluations that they themselves deem necessary. There is quite an active market for consultancy firms specialising in this type of evaluations (Speer, *op. cit.*).

Public reaction to evaluation results is greater in cases of education than those of employment policy, where discourse is restricted to groups of experts and practitioners, with almost no coverage in the media.

Furthermore, the evaluation communities in both cases are very different and clearly separated: psychologists, school principals and teachers in the case of education; academic researchers and economists in particular in the case of employment policies. In the words of Speer:

"In Germany, where the institutionalization of evaluation on the national level is relatively lower, the evaluation cultures are very unique to each sector...Although an active national evaluation society exists in Germany, the communities carrying out the evaluations remain fragmented and there is a lack of interdisciplinary discourse on the subject." (Speer, *op. cit.*, pos. 1787)

Evaluation in local government

In Germany there are also no significant studies on evaluation practice and its institutionalisation at local level. There are, however, studies describing the adoption and evolution of performance management and performance monitoring systems introduced under the New Steering Model in the 1990s.

The situation of such practices at the municipal level in Germany is a curious mixture of the elements found at the same level in Sweden, France and the United Kingdom. In Sweden and France, the emergence of such practices followed an bottom-up dynamic and become widespread, according to Kuhlmann, over the course of the 1990s, when around 82% of local governments in Germany adopted reforms of this type. In the case of France, systems of indicators that focussed on inputs, activities and results were adopted in the majority of cases.

Inter-municipal comparison mechanisms were also put in place (comparing the cost and quality of services, among other aspects) in 60% of German municipalities. The resulting system is similar to the British system in terms of its complexity and seems overly bureaucratic in some cases:

"There is a precarious tendency of adding to the already existing bureaucracy of legal rules and procedures a new one that consists of indicators and 'products'. This 'bureaucratic outgrowth' of performance indicators in Germany is illustrated by the management chart for the citizen services authority (Bürgerservice). The KGSt proposes 66 performance indicators with detailed measurement instructions." (Kuhlmann, op. cit. page 339)

Moreover, German local authorities are free to decide whether or not to publish the information collected through these mechanisms. Inter-municipal comparisons are not typically published, and the information produced is hardly used:

"In practice, performance management instruments are barely used for political or administrative decision-making. In most cases, the information about the products and performance is neither applied for budget approval nor for the reorganization of administrative procedures or inter-municipal comparison. Most strikingly, 14 percent of local authorities that deal with performance indicators do not use this information at all." (Kuhlmann, 2010, page 339-340)

The evaluation system in Germany

General assessment

The evolution of evaluation in Germany follows a pattern similar to that seen in the previous three countries until the mid 2000s: an early emergence, promotion during the golden interwar period, and a change of direction in the 1980s and 90s under the rubric of the New Steering Model in favour of performance measurement and monitoring methods. The year 2000 saw the beginning of an institutionalisation phase that was more procedural than formal, although it does include regulation through legislative clauses that impose evaluation obligations in certain areas. The country's federal structure and the high degree of autonomy enjoyed by ministries favoured diversified implementation of evaluation, with a greater presence in certain public policy sectors, as in the case of education and of employment policies. In Germany, evaluation is normally conducted externally to the administration by research institutes, either independent or organically linked to the bodies responsible for policy. In any case, evaluation is conducted in accordance with high methodological standards and full autonomy to execute and publish the studies. There is a high level of compartmentalisation between academic disciplines, as well as between academia and the administration. Unlike in the United Kingdom and the Netherlands, the German Court of Audit has a limited influence in this area. The use of evaluations for the purpose of improving policies seems high at the federal level.

5. Evaluation in France

Institutionalisation of evaluation begins in France in the 1980s under the administrative reform initiatives promoted by successive socialist governments under the presidency of François Mitterrand (1981-95). After the failure of the *Rationalisation des Choix Budgétaires* (RCB)²⁷ initiative, the *Direction de la Prévision* of the Ministry of Economy, Finance and Budget attempted to transfer the inter-ministerial network created for the RCB experience into an evaluation network in order to benefit from the experience gained in public policy analysis.

For the purpose of launching the initiative and raising awareness, in December 1983, a conference on public policy evaluation was held in Paris, jointly organised by the Ministry of Economy, Finance and Budget and the *École des Hautes Études Commerciales*. From that moment onwards, evaluation was considered in all areas of French public administration, although there was a marked lack of enthusiasm.²⁸

After the conference, the issue of evaluation was revisited by two ministers: Jacques Delors (Economy and Finance) and Michel Rocard (Plan), who commissioned a report from Michel Deleau on the need for the institutionalisation of evaluation practice. A working group comprised of external experts participated in the study, which culminated in presentation of the final report *Évaluer les politiques publiques: méthodes, déontologies, organisation: rapport* in 1986. Shortly after the presentation of the report, there was a change of government, with Chirac taking office as Prime Minister and Mitterrand as the President of the Republic (first period of cohabitation). The new Minister of Finance, Edouard Balladur, refrained from applying and following the recommendations of the report.

In 1988, with Michel Rocard as Prime Minister, modernisation of public services became linked to evaluation, which was seen as an essential counterbalance to autonomy. To this end, a new external report was commissioned under the title '*L'Évaluation des politiques publiques et des actions publiques*'²⁹, which would serve as a basis for the report entitled '*Renouveau du service public*' of 23 February 1989. This latter report placed an emphasis on the democratic control approach, defining evaluation as a judgement reached by democratically elected representatives on the value of public policies (Jacob, *op. cit.*; Fontaine and Monnier, 2002). This therefore stressed the need to increase evaluation sources in order to strengthen the balance of powers, especially

27. French adaptation of the Planning, Programming and Budgeting System (PPBS), introduced in the US Federal Government under President Jimmy Carter in 1979.

28. "Le Commissariat général du Plan n'exprime qu'une tiède bienveillance. Plus incisive, la Direction du Budget fait savoir à la DP [Direction de la prévision] qu'elle n'est pas favorable à une relance du dispositif RCB, même sur une base évaluative. Pour elle, l'échec de la RCB montre que les politiques publiques se gèrent à coup de rapports de force où la raison intervient peu." (Spenlehauer, cited by Jacob, *op. cit.* page 52).

29. Viveret, P., *L'évaluation des politiques publiques et des actions publiques*; followed by *Propositions en vue de l'évaluation du revenu minimum d'insertion : rapports*, Paris, la Documentation française, 1989.

in relation to the executive branch, to prevent it from monopolising expert knowledge on public interventions.

The inter-ministerial evaluation mechanism derived from the aforementioned report represents the first institutional framework for evaluation in France. It is comprised of four different bodies: *Commissariat général du Plan*³⁰; *Fonds national du développement de l'évaluation*; el *Comité interministériel de l'évaluation* (CIME); and *Conseil scientifique de l'évaluation*. The evaluations promoted within this framework, and particularly by the *Conseil scientifique de l'évaluation*, were pluralistic in nature, having been commissioned from evaluation committees comprised of elected representatives, public sector managers and beneficiary groups, who were called on to reach a value judgement on the policies evaluated (Fontaine and Monnier, 2002).

Numerous evaluations are conducted in this manner, but their recommendations were not followed (Jacob, *op. cit.*). In addition, over time, the mechanism was increasingly seen as being too rigid by senior officials and researchers alike. In general, the regulation of evaluation tasks was seen to be excessively rigorous and scientifically complex, with almost no public officials trained to use the mechanism³¹.

The progressive lack of interest among public sector decision-makers in the work of the Inter-Ministerial Evaluation Committee (CIME) is apparent in the lack of committee meetings after March 1993, the absence of commissioned evaluations, and the committee's failure to reappoint its members at the end of their mandates. Finally, in late 1998, the evaluation mechanism in the executive branch was amended, making the *Conseil national de l'évaluation*, and the *Commissariat général du Plan* responsible for the evaluation of public policy at both the state and local level. The *Conseil national de l'évaluation* consisted of fourteen members, each appointed for a three-year term and who could be representatives of national organisations, local organisations or experts in evaluation. One of the primary duties of the *Conseil* was to promote an annual evaluation programme and provide technical assistance and training for its implementation by the public bodies responsible for the policies to be evaluated. In addition, it was charged with overseeing the quality of evaluations and disseminating the results. Nevertheless, the new inter-departmental mechanisms also met with little success (Jacob, *op. cit.*).

Despite the successive failures of inter-departmental evaluation mechanisms, evaluation continues to be practised in certain public intervention sectors, such as education. From 1987 onwards, we see a series of initiatives aimed at institutionalisation of evaluation in the education area, as demonstrated by the creation of the *Direction*

30. Public institution created in 1946, which was in charge of economic planning in France through the definition of five-year plans up until 2006. It subsequently became the Centre d'Analyse Stratégique (CAS).

31. "Quand on pose l'évaluation des politiques publiques dans le sacré, ce n'est pas étonnant qu'on n'en fasse pas." (Audebrand, É., cited by Jacob, *op. cit.*, page 79).

de l'évaluation et de la prospective within the *Ministère de l'Éducation*. In early 2000, the new Minister of Education, Jack Lang, inaugurated the *Haut Conseil pour l'évaluation de l'école*. The end of the 1980s also saw the creation of the National Committee for the Evaluation of Urban Policy, as well as various other evaluation agencies in the fields of scientific research and vocational training.

Of particular significance in the field of social policy was the introduction of the evaluation clause in the Minimum Integration Income Act adopted by the French National Assembly in December 1988. The text of this clause contains a provision that makes continuation of the programme conditional upon the results of its evaluation at the end of a three-year term. However, at the end of the stipulated term, this clause was amended, and the evaluation requirement was dropped. Notwithstanding, a subsequent version, the *Loi de la Revenu de Solidarité Active* approved in late 2008, served as a basis for the creation of an innovative evaluation initiative that is still in place today: the *Fonds d'Expérimentation pour la Jeunesse*³². More specifically, the regulation stipulated that this Fund would receive contributions from the state and from other associated agencies to define, finance and pilot one or more experimental programmes aimed at improving the social and professional integration of young people aged 16 to 25. The Fund financed such programmes, as well the mandatory independent impact assessments associated with them. At the end of 2010, the Fund's purpose and target population was further extended to favour the educational success of students and improve the social and professional integration of young people under the age of 25 (Gurgand and Valdenaire, 2012). This is an interesting initiative framed within a new trend for social innovation underpinned by rigorous impact assessments. But its interest also lies in the financing and the mixed public-private management model of the Fund (Lázaro, 2012b).

As regards the role of the French parliament, it initially set up an agency shared by the National Assembly and the Senate: in 1995 a draft law proposed the creation of a parliamentary office of public policy evaluation (OPEPP) to be responsible for conducting both prospective and retrospective evaluations of the suitability of legal, administrative or financial resources devoted to public policy with regard to the observed or expected outcomes. The office consisted of two branches (a National Assembly branch and a Senate branch) and received a range of criticism for the fact that its work overlapped with that of parliamentary commissions and because its internal expertise was seen to be isolated from the decision-making process. Ultimately, this first parliamentary experience in creating its own public policy evaluation mechanism ended in failure and the office was dissolved in 2001.

Afterwards, each chamber established its own evaluation agency. The Senate, which wanted to lend continuity to the work of OPEPP, created the *Comité d'évaluation des*

32. <http://www.experimentation.jeunes.gouv.fr/>

politiques publiques. In the initial stages it consisted of 16 members of the Senate who had previously been involved in the work of the OPEPP. At a later stage, the committee became attached to the Finance Commission. The evaluation studies promoted by the Commission were conducted by external experts and required the appearance of the administrative managers responsible for the evaluated programmes.

For its part, the National Assembly opted for a different mechanism, la *Mission d'évaluation et de contrôle* (MEC), which was created following a working group on the effectiveness of public spending and parliamentary oversight. The purpose of the MEC is to exercise the prerogative powers of parliament in relation to controlling the use of public funds, as set out in the Budget Act. In February 1999, in order to speed up the creation of this body, it was not created as a permanent committee but rather as one which must be reapproved every five years by the Finance Commission. This procedure has the advantage of obliging Assembly members to express their will to continue with the work on a yearly basis, which is important in terms of fostering an explicit commitment to evaluation. Furthermore, rather than merely commissioning new evaluations, the MEC analyses existing evaluations and publishes their contents, acting as a kind of sounding board, although with a relatively limited level of success³³.

Today the National Assembly has three evaluation mechanisms:

- The *Mission d'évaluation et de contrôle* (MEC), which has been up and running since 1999 and fulfils a role inspired by the British *National Audit Office* of overseeing the effectiveness and efficiency of public spending, with the ability to conduct in-depth evaluations of sectoral policies.
- The *Mission d'évaluation et de contrôle des lois de financement de la sécurité sociale* (MECSS), created in late 2004 to oversee the use of Social Security funds through audits, studies, etc.
- The *Comité d'évaluation et de contrôle* (CEC), created in May 2009, with express responsibility for the evaluation of public policies in areas affecting more than one parliamentary committee.

In contrast, the Senate does not seem to have any specific evaluation body, although it does have a specific evaluation and forecasting programme³⁴ structured around different topics, which are assigned to different commissions.

As regards the *Cour des comptes*, in France the role of this organisation with respect to evaluation is ambiguous. Unlike the very active role of its counterparts in Sweden and the Netherlands, in France, on one hand, it pronounces itself in favour of evaluation

33. "Les suites accordées aux travaux de la MEC sont très limitées et il arrive même que le Gouvernement prenne le contre-pied des suggestions qu'elle formule." (Jacob, *op. cit.*, page 76).

34. Programme de contrôle, d'information, d'évaluation et de prospective du Sénat. <http://www.senat.fr/contrôle/2014/organismes.html>

activities, however, in practice it acts in a rather more prudent manner (Jacob, *op. cit.*), avoiding issuing opinions on decisions adopted by the executive and legislative branches and instead focussing on analysing the consistency between these decisions and the actions implemented. However, since the constitutional reform of 2008³⁵, the *Cour des comptes* has been collaborating directly with the *Mission d'évaluation et de contrôle (MEC)* of the National Assembly.

Evaluation in regional and local governments

March 1982 saw the enactment of the Rights and Freedoms of Municipalities, Departments and Regions Act, which established the decentralisation of central government powers to these local administrative bodies. Among other things, the new law stipulated that public service contracts between the state and the regions should be formalised. This obligation, coupled with those derived from the management of EU structural funds, in turn led to the proliferation of evaluation in regional administrations, alongside the creation of specific organisational units dedicated to evaluation. However, as in the case of the municipalities, the methods of institutionalisation and the evaluation practices adopted varied widely. For example, in 1991 the Rhône-Alpes Region established a Regional Evaluation Committee to serve as a forum for deliberating the actions to be evaluated and analysing the respective results, and a scientific committee for ensuring that pertinent methodological corrections were made. In contrast, in Brittany, similar bodies were created but with state participation. Throughout the 1990s, evaluation mechanisms were gradually extended to different regions (Midi-Pyrénées in 1992, Pays-de-la-Loire in 1993, Lorraine in 1994, etc.).

As regards municipalities, the central administration also failed to implement a specific and consistent evaluation mechanism. However, the municipalities themselves showed particularly high levels of activity (Kuhlmann, *op. cit.*). At this point we should note that local administration is highly fragmented in France, with some 36,000 municipalities, 75% of which have less than 1,000 inhabitants. Consequently, the progress of evaluation has been somewhat irregular, with advances made at 'two speeds': at a faster pace in larger cities, inter-municipal cooperation agencies and departments; and at a slower pace or not at all in medium-sized or small municipalities. In any case, this has been a voluntary, "bottom-up" and varied process, focussing mainly

35. The *Comité de réflexion et de proposition sur la modernisation et le rééquilibrage des institutions*, was created in July 2007, chaired by Édouard Balladur. The work of this committee would, among other things, lead to the creation of a draft law later approved by the National Assembly in July 2008 which would redefine the role of parliament in the control of the executive branch. More specifically, the law (*loi constitutionnelle no 2008-724 du 23 juillet 2008 de modernisation des institutions de la Ve République*) amends Title V of the Constitution, introducing a new article, 47.2, which establishes the following: "*La Cour des comptes assiste le Parlement dans le contrôle de l'action du gouvernement. Elle assiste le Parlement et le gouvernement dans le contrôle de l'exécution des lois de finances et de l'application des lois de financement de la sécurité sociale ainsi que dans l'évaluation des politiques publiques. Elle contribue à l'information des citoyens via des rapports publics.*"

on monitoring systems rather than on evaluation, and adapted to the local context and to specific services, although centred on easily measured indicators, inputs, activities and results, with hardly any impact indicators at all. The local adoption of these policies favours their internal use but makes inter-municipal comparisons more difficult. In addition, there is generally a resistance towards public dissemination of results (Kuhlmann, *op. cit.*).

Professionalisation of evaluation

The *Société française de l'évaluation* (SFE) was formally founded at its first conference held in Marseille in June 1999. The role of the SFE is one of education and promoting evaluation. It holds an annual conference which brings together members of the administration, academia and the private consultancy sector. The SFE has created various different working groups on a wide range of evaluation aspects (methods, practices and capacity for evaluation, etc.), as well as regional sections.

The evaluation system in France

General assessment

The development of evaluation in France begins in the 1980s in the form of successive attempts at formal institutionalisation in both the executive and the legislative branches which, after various trials and failures, resulted in a pluralistic and decentralised evaluation practice. Evaluation occupies a prominent place in the political discourse linked to the administrative reform of the 1980s, although attempts at institutionalisation failed to jell in those years. The first evaluation mechanism was implemented in the executive branch in 1990s, was inter-ministerial in nature and rested on four pillars: the *Commissariat général du Plan*, the *Fonds national du développement de l'évaluation*, the *Comité interministériel de l'évaluation* (CIME) and the *Conseil scientifique de l'évaluation*. Numerous evaluations were conducted, but their recommendations were not used and the mechanism itself was considered too rigid. Finally, in late 1998 the mechanism was simplified to include only the *Conseil national de l'évaluation* and the *Commissariat général du Plan*. The *Conseil national de l'évaluation* was comprised of members of national and local bodies and experts in evaluation. Its role was to promote an annual evaluation programme and to provide the technical assistance and training required for its implementation by public institutions. In addition, it was charged with overseeing the quality of evaluations and disseminating the results. However, the new inter-departmental mechanisms also met with little success. Despite the successive failures of inter-departmental evaluation mechanisms, evaluation continues to be conducted within certain areas of intervention such as education, social policy (after the entry into force of an evaluation clause in the Minimum Integration Income Act), and urban policy, and includes rigorous impact evaluations. Furthermore, in the legislative branch, the National Assembly has various different evaluation agencies, although their effects in terms of control over the executive branch seem limited, as is the role of the *Cour des Comptes*.

6. Evaluation in Belgium

In Belgium, as Varone, Jacob and De Winter (2005) state, the evaluation of public policy still has a fairly weak presence, both in institutional terms and in terms of the existence of a favourable culture. This situation is partly due to the distinctive characteristics of the political system and particularly to the weakness of the parliamentary system and the central role of political parties in defining the political agenda. Both of these factors, among others, hamper the functioning of accountability mechanisms and the institutionalisation of evaluation.

The fragmentation of Belgian society for religious, socio-economic and cultural reasons translates into a multi-party political system and the habitual formation of coalition governments. It is an example of a consociational democracy, with deep divisions between the different sub-cultures. Coalition governments maintain their fragile equilibrium by minimising interaction with political agents outside the coalition. The process of creating policy is 'captured' by ministerial cabinets comprised of trusted party members (Varone, Jacob and de Winter, 2005).

This situation is detrimental to evaluation in two senses. Firstly, it is not conducive to the introduction of technical factors that might contradict the party line and put the stability of the governing coalition at risk. Moreover, the government's control over legislative initiative minimises the chances of any introduction of provisions that favour independent evaluation. Secondly, although the administration is responsible for the implementation of policies, it does not have the power or resources itself to promote the evaluation of these policies, and even if they were evaluated, the chances of the results being used and disseminated are remote. In this context, promotion of evaluation seems to directly depend on the will of the parties in government. That said, promotion of evaluation is also largely absent from electoral campaigns and programmes (Varone, Jacob and de Winter, 2005).

Some progress has been made formally, although in Belgium the administrative reforms inspired by New Public Management have largely forgotten evaluation (Jacob, *op.cit.*). As such, despite some isolated initiatives, public policy evaluation continues to be rare, scarcely systematic, and dispersed across the Belgian administration and government, mostly due to a lack of awareness of its usefulness.

Initially, evaluation was used to seek consensus with regard to certain political issues (Jacob, *op. cit.*). One landmark example of this was the processing of the 1990 draft law on termination of pregnancy, where evaluation provided a 'guarantee' in the eyes of the opposition, serving as an instrument of control over implementation of the law and as a tool for warning of possible deviations. Thus, said law, enacted in August 1990, led to the creation of an Evaluation Committee.

That same approach was adopted in 1999 for the law on Belgian international cooperation enacted in May of that year. Said regulation established a mandatory requirement for annual evaluation reports to the parliament, and an evaluation service was created within the Directorate-General for International Cooperation. Another example is the creation, in 2002, of the Federal Control and Evaluation Committee for the law regulating decriminalisation of euthanasia under certain circumstances. As Jacob states, in these cases

"Le recours à l'évaluation apparaît comme un instrument "d'apaisement" au cours de débat éthiques." (Jacob, op. cit, page 71)

In 1999, following the creation of a 'rainbow' coalition of liberals, socialists and ecologists, a government statement placed an emphasis on the evaluation of legislation:

*"Le Gouvernement, avec le Parlement et en collaboration avec le bureau de coordination du Conseil d'État, procédera à une évaluation de la législation. Celle-ci est un moyen d'améliorer l'efficacité, l'adhésion citoyenne et la qualité juridique de la législation. Cette évaluation de la législation permettra de juger leur fonctionnement et leur efficacité, de sorte qu'elles puissent être revues et, au besoin, abrogées."*³⁶

However, ultimately, the institutionalisation of evaluation did not form part of the agenda of the Verhofstadt Government in that first stage between 1999 and 2003.

As for organisational agencies with evaluation responsibilities, there are two that are worthy of note. On one hand, within the executive branch, is the *Bureau fédéral du Plan* which, following a legislative reform amending its duties in 1994, became responsible for analysing and forecasting socio-economic progress and its determining factors, and for evaluating the consequences of decisions on economic policy with a view to improving the rationality, effectiveness and transparency of such policies³⁷. This office falls under the authority of the Prime Minister and the Minister of Economy, and its lines of action are approved by the Council of Ministers. It is important to note that, unlike in other countries such as the Netherlands, France or Switzerland, in Belgium there was no attempt to create cross-cutting mechanisms for the institutionalisation of evaluation in the executive branch (such as inter-ministerial committees or similar bodies).

In 2001, the *Association universitaire de recherche sur l'action publique* of the *Université Catholique de Louvain* conducted a study on evaluation practice in the Belgian administration. The study showed that most public sector bodies used evaluation as a

36. « La voie vers le XXI^e siècle ». *Déclaration gouvernementale prononcée devant le Parlement le 14 juillet 1999*, Bruxelles, SFI, 1999, P.22. (Cited by Jacob, op. cit. page 51.

37. Law of 21 December 1994, *Moniteur belge*, 23 December 1994.

management tool. Evaluations are conducted at the discretion of the pertinent administrative bodies, often with no express political mandate, and are executed and used internally. Neither political decision-makers nor beneficiaries of policies are generally involved in the evaluations.

Moreover, in 1998 the power of the Court of Audit (*Cour des Comptes*), which is attached to parliament, were extended to cover the oversight of the effective and efficient use of public resources³⁸. Unfortunately, the evaluations conducted since then have had little resonance. A shortage of qualified evaluators, certain limitations in relation to technical quality and access to information, a lack of stakeholder involvement in evaluations, and the absence of effective practices for disseminating results—coupled with the aforementioned institutional factors—may have contributed to this situation.

As for parliament, generally in Belgium, as in the Netherlands, the legislative branch does not have a leading role in the development of evaluation (Jacob, *op.cit.*) other than the approval of the specific legislative provisions on evaluation mentioned above. It does not create any *ad hoc* internal agencies specialising in evaluation.

Finally, although some authors (Varone, Jacob, de Winter, 2005) suggest that the state's federal structure may act as an additional barrier to evaluation (due to the greater level of fragmentation and the complexity of policy creation and implementation processes, and due to conflicts of powers that hamper coordination and transparency, etc.), they do recognise that it could also work in favour of evaluation since it provides the ideal scenario for a 'laboratory' for territorial experimentation and comparative programme evaluation. They also believe that the example provided by the Wallonia region—with more advanced evaluation practices, forged as a result of the obligations associated with EU structural funds—may serve to spur the rest of the country.

The implementation of evaluation in the region is gradual, with an emphasis on establishing a favourable culture, supported by the *Société Wallonne d'évaluation* (SWE), created in 1999 and the *Institut wallon de l'évaluation, de la prospective et de la statistique*³⁹ (IWEPS), set up in 2004. The IWEPS was created with the mission of becoming an independent research centre, producing knowledge and providing strategic guidance to regional government institutions and to social and economic agents. The organisation recently underwent a restructuring process, in 2013, which saw the creation of a research and evaluation division, an information and data division, and a forecasting division.

38. The role of the Belgian *Cour des comptes* in relation to evaluation is defined in the law of 10 March 1998, which amends the previous regulation and stipulates that the Court must control the effective use of public funds.

39. <http://www.iweps.be/>

The evaluation system in Belgium

General assessment

Belgium is a federal state divided into different linguistic and cultural communities, a division that gives its institutions a high degree of complexity, with their organisation and functioning conditioned by the need to represent all of the different communities. That representation is channelled through political parties which permeate all government institutions. The centrality of the political parties and the weakness of the parliament in its role as the body exercising control over the executive branch have hampered the functioning of accountability mechanisms and the institutionalisation of evaluation. This situation is not conducive to the introduction of technical factors that might contradict the party line and put the stability of the governing coalition at risk. Moreover, the government's control over legislative initiative minimises the chances of any introduction of provisions that favour independent evaluation. In this context, promotion of evaluation seems to depend directly on the will of the parties in government. That said, it is predominantly absent from electoral campaigns and programmes.

Therefore, public policy evaluation continues to be rare, scarcely systematic, and dispersed across government and the Belgian central administration, mainly due to a lack of awareness of its utility. However, at sub-state level, evaluation mechanisms have been under development since the mid 2000s, with the Wallonia region at the forefront with the creation of the IWEPS (Wallonia Institute of Evaluation, Forecasting and Statistics). Nonetheless, there is little information on the current functioning of the system at either the federal or state level to allow for a more detailed analysis.

7. Evaluation in Southern European countries: Italy and Spain

The countries of Southern Europe have attempted to adopt Napoleonic political-administrative traditions, although not entirely successfully (Ongaro, 2009; Pollitt and Buckaert, 2004). Both the inherent aspects of that tradition and the contextual, political and social factors that have prevented the complete introduction of evaluation in Southern Europe, have generally proved to be non-conducive to its development.

In the Napoleonic model of public administration, the approach to the role of the administration is still predominantly a legalistic one (as opposed to managerialistic or discretionary); the values that guide the provision of public services favour uniformity (rather than encouraging difference and experimentation); and social groups are perceived as non-legitimate participants in the public policy process. In fact, the government constitutes the highest authority for matters of administration and control, which is exercised through a bureaucratic system governed by administrative law. For its part, the parliament creates parliamentary commissions to investigate certain issues, but there is no true parliamentary tradition of executive oversight. The Courts of Audit are judicial bodies that differ greatly from the audit offices found in Anglo-Saxon or Scandinavian countries.

However, these countries do share the typical inflexibilities of this model associated with the inherent weaknesses of a social and political culture that has not yet managed to eradicate the presence and influence of patronage networks among political parties, the administration, and the social and economic sector (Ongaro, 2009; Kickert, 2011). This has resulted in the excessive political penetration of the administrative system, an issue that has been identified in the existing literature as one of the main factors responsible for the failure of attempts to modernise administrations in Southern Europe (Kickert, 2011).

This is the case because this excessive politicisation simultaneously affects three key dimensions of the functioning of administrative systems and, by extension, the policies they attempt to promote. Firstly, it accentuates the aforementioned patronage, which favours particular interests over general interests. Secondly, it leads to the squandering of highly qualified human capital and de-incentivises professional development and management of human resources. Finally, it introduces endemic discontinuity in public policy, since each government tends to impose a new agenda, immediately rejecting or substantially changing any initiatives that originate with previous teams.

Furthermore, public sector unions protect the civil service function and public sector employment in general as it is today, with all its faults and inefficiencies, and have stated their express opposition to evaluations of the performance of civil service staff working within the administration. There is also still a high level of tolerance among the public of deficiencies or poor performance in the delivery of public services,

although the cuts resulting from the current crisis have served as a catalyst in some cases.

All of this has created a context that does not bode well for the development of administrative reform processes. Any progress made in these areas has been limited and fragmented, taking the form of discontinuous processes usually focussed on specific organisations rather than major cross-cutting reforms.

Evaluation tended to be introduced in this environment from outside, in reaction to the obligations derived from the management of EU structural funds. In later years, we have the emergence of internal processes aimed at the formal institutionalisation of evaluation, but these processes have not led to the effective implementation and operation of authentic evaluation systems, at least for the time being. However, as we saw in the case of Belgium, in both Italy and Spain these processes exist alongside certain sub-state initiatives which have a more flexible approach and show interesting signs of evolution.

7.1. Evaluation in Italy

Evaluation was introduced in Italy a result of external factors, with attempts at its implementation within a political and administrative context that was unfamiliar with such practices.

According to Stame (2012), in this country, the penetration of political parties in all state institutions, including the public administration, and the consequent patronage and instrumental use of the administration for partisan interests, has led to a traditional mistrust and estrangement between the Italian public and the state. This estrangement has prevented the development of external public and social pressure of sufficient force on the administration in favour of increased effectiveness, efficiency and transparency. Moreover, the state is perceived as an enemy, often with an excessive zeal for control and repression, which heightens the feeling of mistrust and public alienation.

However, there is a vital and well organised civil society that has contributed to the existence of efficient administrations at local and community level, as well as various NGOs and community initiatives that provide services to those in need. The role of the organised civil society has become increasingly important due to the recent fiscal crisis and public sector budget cuts.

In this context, Ongaro and Valotti (2008, cited by Stame, 2012) characterise the Italian political-administrative system as a discordant combination of centralised institutions⁴⁰,

40. Nonetheless, although the political system is centralised, regions and municipalities enjoy a high level of

tolerance for inefficiency, and democratic values. This lack of alignment gives rise to lively political debate and is the context in which various administrative reforms have been carried out since the 1990s.

Practices

The life cycle of public policies is contemplated from a legalistic standpoint, and the approach to implementation is predominantly regulatory. All attention is focussed on the legislative process. Once the relevant regulation is approved, implementation is understood to be practically automatic, with no consideration given to possible alternatives, etc. In general, little importance is placed on the implementation process and the actions of the administration in general (partly as a result of the vicious cycle of mistrust, estrangement and ignorance to which we referred earlier), and therefore little effort is made to improve performance. All attention, as we said, is focussed on the legislative process, which effectively means attempting to resolve problems of inefficiency and poor performance by enacting new laws. In the words of Stame:

"...all the attention is turned on the legislative function, while the executive function (performed by the public administration) is considered less important. So, if it does not take place as expected, the tendency is not to address administrative problems but to promulgate a new law." (Stame, op. cit., pos. 504)

Public Administration Reforms

The early 1990s saw ratification of a series of laws aimed at reforming the public administration, known as the 'Bassanini laws' after the centre-left government minister who promoted them. These initiatives included both 'traditional' public sector reforms (Wollmann, cited by Stame, 2013) such as strengthening the leadership of the Prime Minister and Mayors; decentralising powers to the regions; increasing transparency, democratisation and citizen participation; and other reforms inspired by New Public Management, such as outsourcing the provision of services, agencification, observance of employment law in the hiring of public-sector managers and civil servants subject, etc.

One reform of particular importance was that aimed at clearly demarcating areas of political action and administrative action, or public management. This resulted in the hiring of public sector managers from outside the administration through the use of contracts of limited duration subject to private law which also involve the establishment of objectives and monitoring and evaluation of levels of attainment. This in turn provided a greater level of autonomy to these managers and led to the creation of evaluation units in all administrations (aimed at evaluating or measuring performance).

autonomy, and the constitutional reform of 1999-2001 introduced a 'quasi-federal' system (the state, the regions and the municipalities are placed at the same level as the institutions of the Republic).

However, these reforms have met with limited success: politicians have continued to have a decisive influence on the appointment of public sector managers, who are removed from their posts when politicians change. Bassanini offers an eloquent and sobering summary:

"Accountability, merit, performance assessment are still the exception, not the rule; they have been overwhelmed by the spoils system, by the scattershot distribution of performance bonuses, by the explosion of the costs of politics, by the resistance of an entrenched bureaucratic culture, by the conservatism of the bodies responsible for auditing, by the inability of the political class to conceive of public policies in terms of strategies, objectives and quantifiable, measurable results." (Bassanini, 2009, page 375)

Specifically in relation to performance evaluation, the same author indicates that the main reforms, contained in Legislative Decree 80/1998, established that

"...for each administration the political authorities are responsible for defining public policies, translating them into strategic guidelines, setting precise, quantified goals, and introducing objective, reliable systems of evaluation; that administrators are completely autonomous and accountable for managing their units; that careers, promotions, removal from position, and even a part of pay must be related to results; and that public employees can be dismissed under the same rules that prevail in private law."

In relation to the results:

"...only a few administrations have applied them with conviction. The objectives, when they exist, are often generic and vague; performance is either not measured at all or is assessed in purely summary and discretionary fashion..." (Bassanini, 2009 pages 379-381)

As regards the main causes of failure, the author cites the reticence of the political classes to set accurate productivity objectives and deploy rigorous and reliable mechanisms for measuring performance. This shows the resistance of the political classes to doing away with the innate and wide-ranging discretion of a culture of patronage. Furthermore, setting relevant and viable objectives is not easy; it requires time, analysis and reflection by individuals holding political office and professionals within the administration, tasks that are far from normal practice in both cases.

Another contribution to this failure was the lack of continuity of reform policies in successive governments, despite it being a reform agreed between the centre-left government and the centre-right opposition (the reform which began in the 1990s). The aforementioned failed reform towards a federal system⁴¹ was also a contributing factor.

41. According to Bassanini (op.cit), this failed due to the absence of a supremacy clause in the constitution, in

An additional problem was the lack of resources allocated from the very start, since the state of the public finances prevented the investments needed for the reforms to go ahead (on the contrary, one priority objective was to reduce public spending), including computerisation, recognition of merit and good results, improving the quality of public services, staff training, hiring of young personnel and/or experts in new areas prioritised in the reform, etc. (Bassanini, op. cit.).

The administrative reform included the creation of evaluation units in different public bodies. However, in numerous cases, these units were established in a somewhat hasty manner, as though they were just another administrative unit, replicating the control functions of previous units rather than considering what their objectives should be and how to achieve them (and in particular what work methods should be adopted and what professional profiles would be needed to develop and implement them) (Stame, 2012).

Given that these issues had not been considered, the evaluation units created were staffed with generalists from a range of disciplines (lawyers, accountants, statisticians, etc.) and/or specialists on certain sectoral policies (health, environment, etc.), without requiring experience in evaluation. As to be expected, the way in which these units operated does not seem to have contributed to the advancement of evaluation.

In contrast, a great effort was made academically (Borgonovi, 2005, cited by Stame, 2013) to establish a system for measuring the productivity of public administrations based on batteries of performance indicators, as well as indicators of inputs, outputs, outcomes, etc. But these indicators were designed without the involvement of the managers and technicians who would use them, which led to significant limitations in terms of the design and actual value of the indicators, as well as their adoption—which became ritualistic—by the administration:

"Unless the personnel are involved in their definition and in interpreting their results, indicators alone can favour goal displacement (typical of public sector ritualism); alternatively, if they measure a failure they may provoke disenchantment instead of a desire for understanding what specific conditions and human capacity may make a program succeed."
(Stame, op. cit., pos. 573)

Evaluation functions have therefore been seen as control functions, and public bodies have developed strategies to evade its influence, which has been counter-productive for the advancement of evaluation:

favour of central government, which has led to legislative duplications between the state and the regions, and has prevented the elimination from state legislation of matters that should be governed by the regions because of the inability to guarantee the creation of the necessary regional regulations in a sufficiently consistent manner so as to safeguard the general interests of the nation and of its citizens throughout the state's territory.

“...the implication was that since administrative personnel are familiar with control processes, they could better understand evaluation if it were presented as a further control; in reality, people are used not only to controls, but also to avoidance of them. This attitude has had negative consequences for evaluation.” (Stame op. cit., pos. 555).

The late 2000s saw the start of a new reform, driven by Minister Renato Brunetta, once again focussed on increasing the productivity of public administrations whilst improving efficiency and transparency. The reform revolved around the concept of merit on the part of executives, managers and all personnel in general within public administrations, and its evaluation through a system of indicators that placed an emphasis on the satisfaction of the users of the services. However, this new attempt at reform also failed, with one of the reasons being difficulty in setting objectives:

“Setting challenging but credible objectives that can trigger significant improvements in the quality of services and efficient performance is a difficult task, requiring, upstream, an ability to elaborate and define the public policies for which the setting of objectives is an essential instrument, and a capacity for dialogue with the administrative units and their executives on the one hand and citizens/customers/users on the other. Without an adequate definition of the objectives, the whole reform mechanism will go nowhere...” (Bassanini, op. cit. page 389)

Within the framework of the Brunetta reform, the *Commissione per la valutazione, la trasparenza e l'integrità delle amministrazioni pubbliche* (CIVIT) was created, which reported to the parliament⁴². The CIVIT was established by Legislative Decree 150/2009, with the role of coordinating and supervising the independent exercise of evaluation functions within public administrations, and guaranteeing the transparency of these activities. In 2014 the CIVIT was absorbed by the *Autorità Nazionale Anticorruzione* before it had conducted any important evaluation activities.

EU structural funds evaluation

The evaluation obligations associated with the receipt of EU structural funds led to the creation of evaluation units (*Nuclei di Valutazioni*) in all regions and ministries with responsibility for the management of such funds, as well as an evaluation network (*Rete dei Nuclei*, created by law 144/1999). These evaluation units were normally located within regional government programme departments and ministries, and were staffed with existing civil servants or experts recruited on an ad-hoc basis, normally with backgrounds in statistics or economics. Evaluations were typically commissioned from external experts, whilst the units themselves managed the respective commissions and contracts.

42. <http://bit.ly/1vTpcVv>

Given the composition of the evaluations units, the studies carried out were not evaluations per se, but rather economic analyses of the impact of the funds and of the executed programmes on the economy of the region and the country. These units are also responsible for providing the Directorate General for Regional Policy of the European Commission (DG Regio) with a specific set of indicators, mostly financial, with no room for analysis of the results obtained or how they have been achieved. In fact, that system of indicators has been implemented in a centralised manner in these units, without contact with the programmes' personnel or beneficiaries. The DG Regio itself has been quite critical of the evaluations conducted. Some regions such as Campania have been capable of developing more robust and relevant methods for the purpose of evaluation. For its part, the Public Investment Evaluation Unit (UVAL), attached to the Ministry of Economic Development⁴³, has been developing training actions and activities to improve evaluation capabilities within the scope of the programmes financed by structural funds.

To conclude, the characteristic features of the institutionalisation of evaluation in Italy can be summarised as follows (Stame, op. cit.):

- *Centralism:*

"Evaluation units (institutions) are created from the above, they receive the templates for evaluation from the centre. This influences the evaluation practice: rarely do the evaluation units work with their immediate evaluatee... This produces a resistance from the evaluatee in providing information and participating in the evaluation. Therefore, it is more difficult for the centre to understand why something worked or did not..." (Stame, op. cit., pos. 817).

- *Legalism:* what matters is that evaluation is regulated and that the law is implemented. The value of evaluation for learning, improving and enhancing accountability takes a back seat.
- *Normativism:* as evaluation has been considered as just another administrative activity, it has been assumed that there is only one correct way to conduct evaluations and there has therefore been no discussion with regard to the methods used, etc. (although recently there has been more debate in this respect, allowing for the introduction of experimental methods, perhaps as a result of the changes associated with evaluation in certain European Commission programmes). In its most negative expression, this has led to passive acceptance of systems of indicators imposed from above, with no questioning of their relevance or value.

43. http://www.dps.mef.gov.it/uval_eng/presentation.asp

In the words of Stame:

"All this has given rise to an evaluation culture that is characterized by lifeless institutions (evaluations units undistinguishable from existing administrative departments), values of uncertain status (transparency and independence declared but not really pursued) and malpractices of accomplishment (avoiding relevant choices, most notably in methodology)" (Stame, op. cit., pos. 809).

The evaluation system in Italy

General assessment

The introduction of evaluation began in Italy in the 1990s, in response to the package of administrative reforms known as the 'Bassanini Laws', which sought to improve the efficiency of the administration through the introduction of measures such as setting objectives and performance measurement. Although a significant portion of the measures adopted were associated with the management of personnel working in the administration, and therefore with individual performance assessments, the package of reforms also introduced evaluation units in different administrative departments. However, these units never really succeeded. An excessively legalistic and top-down approach with no regard for the specific characteristics and demands of evaluation, or the context in which it was to be introduced, spelled the failure of these reforms. In addition to the aforementioned factors, a lack of continuity in the political support to the package of reforms and insufficient resources also contributed to this failure. The late 2000s saw a new package of reform measures which included evaluation. The *Commissione per la valutazione, la trasparenza e l'integrità delle amministrazioni pubbliche* (CIVIT) was created, which reported to the parliament. Its duties included coordinating and supervising the independent exercise of evaluation functions within public administrations, and guaranteeing the transparency of these activities. However, it does not seem to have carried out any important evaluation activities. At a regional level, the evaluation activities associated with the receipt of structural funds has led to the creation of evaluation units in different regions. However, it seems that even at the regional level, a system in which evaluation is fully established has yet to be developed.

7.2. Public policy evaluation in Spain

Since the enactment of the 1978 Constitution, Spain has adopted a highly decentralised state model, which has been gradually implemented and deepened in successive phases, with its development occupying a large portion of the political agenda, together with the advancement of the welfare state in that same period (Gallego and Subirats, 2012). The reform processes of the *Administración General del Estado* undertaken during that time were limited in scope, fragmented and centred mainly on aspects associated with the rationalisation of structures and internal procedures, as

well as bureaucratic professionalisation of the civil service. This was a process characterised by successive changes, with a cumulative and localised effect normally on specific bodies, rather than major cross-cutting reforms⁴⁴.

The practice of evaluation and its degree of institutionalisation are among the lowest in the OECD (Furubo and Sandahl, 2002, Varone et al. 2005, Viñas, 2009), although some progress was noted in the second half of the last decade. In general, as stated by Feinstein and Zapico-Goñi (2010), more than a coherent evaluation system, in Spain there is a diverse constellation of bodies which operate in different sectors of public intervention and at different levels of government, with responsibilities either for evaluation or for related functions.

The factors that explain this situation include (Viñas, 2009): a lesser degree of development of the welfare state and of the scope of associated public investment; a weak tradition of social research and its application in the study of public policies; the predominance of a legal-administrative culture among the political elite and in the public sector, primarily oriented towards controlling legality at the expense of effectiveness; a dearth of incentives for promoting evaluation, within both political and administrative environments; and a parliamentary system at a state and autonomous communities level that is subject to party discipline, with limited possibilities of effectively controlling the executive branch.

Fundamentally brought about as a direct result of the obligations associated with the receipt of EU structural funds, since the 1990s, the practice of evaluation has been developed, although in a somewhat fragmented and piecemeal manner (Viñas, 2009). In some sectors, such as health policy⁴⁵ (in close contact with scientific practice and culture) and development cooperation policies⁴⁶ (exposed to the generalised evaluation practices of international donors), evaluation has penetrated at a far faster pace. In other sectors such as primary and secondary education, there are monitoring systems in place⁴⁷, and in the higher education sector there are evaluation, certification

44. Such as the reform of the Spanish Tax Administration Agency, the National Social Security Institute, or the Postal Service (Parrado, 2008) or the reform of the Catalan Health Institute in Catalonia (Gallego, 2013). Although with limited results, in the case of the General State Administration, exceptions include the introduction of Programme-based Budgeting in 1984, ratification of the Organic Law on the Organisation and Functions of the General State Administration (LOFAGE) in 1997, the Basic Statute for Public Employees in 2007, and the Agency Act in 2005.

45. For example, see the National Health System Quality Plan <http://bit.ly/1vrJSu5>.

46. See Argilés (2014) for an overview of the evolution of evaluation of international development cooperation policy in Spain. Evaluation in this field was promoted from the mid 1990s and has seen a particular push since 2005, following creation of the Directorate-General for Planning and Evaluation of Development Policies (DGPOLDE). However, that push has not yet translated, in terms of resources dedicated to evaluation, to the production of quality evaluations and utilisation of their results. Since 2008, the institutionalisation process has stagnated somewhat. Nonetheless, the emphasis placed on evaluation in the Fourth Master Plan for Spanish Cooperation (2013-2015), as well as the new evaluation policy derived from said plan and approved in early 2013, seems to mark the beginning of a new, more favourable cycle.

47. We refer to the State System of Education Indicators, which includes indicators of educational results based on standardised tests associated with basic skills, key skills, early-school-leaving rates, etc. This system falls under

and accreditation systems for teaching, faculty and institutions⁴⁸, although none of these systems involve the evaluation of public policies in these areas. In many other sectors, the practice of evaluation is scanty or as-yet non-existent—especially if we are referring to rigorous evaluations of effectiveness.

The Spanish Evaluation Agency

Since 2005, various events at the autonomous community and state level have paved the way for the introduction of evaluation on the political agenda and the will to move towards institutionalisation.

Among these events is the creation of the State Agency for Evaluation of Public Policies and Service Quality (AEVAL), in accordance with the first additional provision of Law 28/2006, of 18 July, on state agencies for improving public services. The agency, which commenced activities in 2007, was created from the existing Directorate-General of Inspection, Evaluation and Service Quality of the Ministry of Public Administrations, absorbing most of its personnel (although lacking experience in public policy evaluation) under a public function regime.

The AEVAL creation process included a preliminary proposals phase, with a committee of experts put together under an initiative of the Ministry of Public Administrations and including experts in various disciplines associated with evaluation and in public administration⁴⁹. This committee conducted an exhaustive analysis of international experience in the field and presented a document of findings and recommendations. Ultimately, the AEVAL was created as a state agency, initially attached the Ministry of Public Administrations and since 2012 attached to the Ministry of Finance and Public Administrations, acting within the framework of management contracts valid for a term of four years which define the objectives and lines of action for the period in question.

The agency's mission is to promote and conduct evaluations of public programmes and policies managed by the General State Administration, although it also has the option to evaluate policies at the autonomous community level by establishing collaboration agreements with autonomous communities. Furthermore, promoting improvements in the quality of public services also forms part of its mission.

As an executive agency attached to the government, although it has its own legal personality and budget, the AEVAL depends on the Council of Ministers both for appointment of its chairperson⁵⁰ and for the approval of the annual evaluation plan.

the remit of the National Education Evaluation Institute, attached to the Ministry of Education, Culture and Sport (<http://www.mecd.gob.es/inee/>)

48. We refer to the system developed by the National Quality and Accreditation Agency (ANECA) (www.aneca.es).

49. The report issued by this committee is available at <http://bit.ly/1EHr9uq>

50. Since its creation eight years ago, the AEVAL has had three chairpersons, two of whom were career civil

Moreover, the AEVAL is governed by a Governing Council comprised of representatives of the Ministry of Finance and Public Administrations (the Secretary of State for Public Administrations and the Secretary of State for Budget and Spending); the Ministry of the Presidency (the Director General for Relations with Parliament); the Ministry of Foreign Affairs and Cooperation (the Under-secretary for the Ministry); and independent experts of renowned standing. In addition, representatives of any autonomous communities that enter into collaboration agreements with the agency may also formally participate in the Governing Council.

The first management contract undertaken between the AEVAL and the Government (2008-2011) established the evaluation of the National Reform Plan (PNR)⁵¹ as the priority line of action, although other programmes could be included in annual evaluation plans at the request of the different ministries. Within this framework, the agency has conducted policy evaluations in various different sectors, with a particular focus on the design and implementation of policies and with a range of different methodological approaches. In 2010, the AEVAL also published a guide on the fundamentals of public policy evaluation, as the first item in a collection yet to be completed.

Since its creation, the public policy evaluation activities of AEVAL have been undermined by a range of factors. One of the most significant factors is undoubtedly the inflexibility of the workforce assigned to the agency, made up of some 60 civil servants, nearly half of whom are responsible for support and administrative management functions within the agency, with the rest divided between the evaluation and service quality departments. The agency's origin as the former Directorate-General of Inspection, Evaluation and Service Quality, coupled with the limited margin it has enjoyed in terms of management of its workforce, has up until now meant that there is a greater level of specialisation in the field of service quality than that of public policy evaluation.

However, the AEVAL has also been forced to conduct its evaluation activities with support from political authorities that is merely formal rather than real. Despite the fact that the creation of the agency was, at the time, a clear expression of the political will to promote evaluation, that will has not been continuous throughout the successive changes of government and ministers. Neither has there been an ability to use this will as a foundation upon which to build a specific policy for the promotion of evaluation within the General State Administration which, beyond merely creating formal evaluation institutions, could also undertake actions for raising awareness, generating capacity within the different public sector departments and ministries, incentivising

servants with technical backgrounds and one with a political background.

51. The PNR, adopted in 2005, is a package of reforms aligned with the Lisbon Strategy which stipulated the full convergence of income, employment and science indicators with the European average by 2010.

evaluation practices, exchanging experiences, and cross-cutting coordination of evaluation activities, etc.

To sum up, between 2007 and 2013 the agency produced 36 evaluation reports. What particularly stands out, and clearly illustrates the point made above regarding the lack of continuity in political support for the promotion of evaluation, is the decreasing number of reports published on the AEVAL website, which went from nine reports published in 2007 to three in 2013, with general elections having had a strong impact (in 2008 when the number of reports published falls by a third, and again in 2012 when no reports at all are published).

Evaluation in the Autonomous Communities

Spain is one of the most highly decentralised European states. As we have seen in the previous sections of this report, highly decentralised political systems offer an interesting context for evaluation, since they constitute a kind of natural laboratory where policies are implemented in some territories but not in others, or where different variants are implemented in different territories, providing opportunities for building counterfactual scenarios that would be much more difficult to create in unitary states.

However, that opportunity has not yet been fully exploited. One noteworthy attempt along these lines took place in November 2009 within the framework of the Inter-administrative State Network for Public Service Quality promoted by AEVAL, and consisted of the creation of a working group on public policy evaluation with the participation of representatives of several Spanish autonomous communities, AEVAL itself, and the Spanish Federation of Municipalities and Provinces (FEMP).

Throughout 2010 and under the coordination of the Catalan Institute for Public Policy Evaluation (Ivàlua), this working group undertook an intensive exchange of information on different aspects of the situation and promotion of evaluation in the participating autonomous communities: Castile and León, Catalonia, Extremadura, Balearic Islands, Navarra, Basque Country, Valencia, as well as AEVAL and the FEMP. However, the 2011 budget cuts (which resulted in the closure the evaluation units recently created in some autonomous communities, staff cuts in others and, in other cases, the mothballing of projects for the implementation of new mechanisms) brought an end the group's activities.

That said, despite the slowdown in the institutionalisation of evaluation in the autonomous communities brought about by the economic crisis, it is important to highlight some of the experiences:

- In late 2005, the regional law on the evaluation of public policies and public service quality was approved in Navarra. This was the first piece of legislation institutionalising the evaluation of public policy in Spain as a cross-cutting government function and

creating a specific body for this purpose: the Regional Commission for the Evaluation of Policy and Service Quality. The Commission is a collegiate body, and its responsibilities include promoting a culture of evaluation within the regional community, defining common criteria and methodologies for the practice of evaluation and creating a public evaluations register. It not only includes members of the different Navarran regional government departments, but also of local agencies and social agents, among others.

The operational management of the system is assigned to the Navarran Institute of Public Administration, within which an Evaluation Service was created with two staff members (including the Service Manager). This service carried out intensive activities to raise awareness, generate resources, provide training, and create inter-administrative evaluation networks within the regional community. However, the arrival of the economic crisis brought the end of the service.

- A little later, in October 2006, the Centre for Research and Public Policy Evaluation was created (currently known as the Catalan Institute for Public Policy Evaluation or Ivàlua) as a public consortium formed by the administrations and universities of Catalonia. The institute commenced activities in 2008 and, despite having a somewhat weak institutional anchorage, it has managed to consolidate its position as the leading evaluation institution in Catalonia. It has done so, primarily, by conducting rigorous impact assessments on active employment, social, and, to lesser extent, education policies, under the external supervision of experts of renowned standing. Furthermore, Ivàlua has developed a strategy for raising awareness and training public sector employees and third-sector employees, which has been warmly welcomed and has enabled significant progress to be made towards building an evaluation culture in Catalonia. However, the lack of a clear vision shared by the political elite (beyond mere discourse) with regard to what evaluation is and which key components of methodological rigour, independence, and transparency should be observed, has hindered its progress.
- Later, in 2010, the Basque Government published the 2011-2013 Public Innovation Plan, which included the promotion of a culture of public policy evaluation among its lines of action. This plan, and particularly the line of action associated with evaluation, was jointly promoted by the Office of the President of the Basque Government (*Lehendakaritza*)—particularly from the Coordination Division—and the Department of Justice and Public Administration through its Innovation and e-Administration Division. However, this line of action seems to have become diluted following the change government in 2012.
- Lastly, in March 2011, the Parliament of the Balearic Islands approved Law 4/2011 on good administration and good governance. The regulation created a Public Evaluation Office reporting to the government through its department responsible for service quality. It gave public-sector agencies in the Balearic Islands the autonomy to conduct the evaluations they deemed necessary, although it also stipulated that

the Governing Council would issue an annual list of priority interventions, whilst assigning priority status from the outset to all interventions with a budget of more than €5 million. That said, both the evaluation office and the evaluation policy contemplated in this piece of legislation are still pending implementation.

To summarise, since the mid 2000s, Spain has made a series of advances towards institutionalisation of evaluation, although only in some cases has this led to an effective evaluation practice. The economic crisis and consequent budget cuts, on top of a lack of a clear political commitment to evaluation, have resulted in a slowdown in the process since 2011.

The evaluation system in Spain

General assessment

Public policy evaluation in public administrations is in its early stages in Spain. It was initially introduced as a result of external requirements associated with the receipt of structural funds. However, it wasn't until the mid-2000s that internal institutionalisation initiatives started to appear. Such initiatives first materialised on the periphery of the state (Navarra), before they were seen in the central administration. In 2006, the Spanish Evaluation and Quality Agency was created with the dual mission of evaluating the public policies of the General State Administration (and those of the autonomous communities if requested) and the quality of public services. The Agency was created from the former Directorate-General of Inspection, Evaluation and Service Quality of the Ministry of Public Administrations, inheriting its workforce of inspectors and management specialists and with hardly any margin for hiring its own expert evaluators. This situation, coupled with the rotation of the agency's Chairmanship, and the lukewarm political support shown after the initial start-up phase, hampered the agency's development. The activities carried out by the AEVAL within the scope of public policy evaluation have mainly centred on the publication of reports, whereas actions for raising awareness and training the personnel of the General State Administration have carried far less weight and do not appear to have been carried out in response to any kind of clear strategy. Furthermore, the promotion of evaluation since 2005 by various autonomous communities including Navarra, Catalonia and the Basque Country, experienced a significant slowdown as a result of the economic crisis, with initiatives being mothballed and, in some cases, mechanisms being eliminated altogether. In Catalonia, the Catalan Institute for Public Policy Evaluation (Ivàlua) continues to operate and has deployed a strategy for raising awareness, providing training and conducting rigorous impact assessments, whilst consolidating its position as a leading institution for evaluation in the autonomous community and in Spain as a whole.

IV. Institutionalisation of evaluation in Latin America

In general terms, and at the risk of overly simplifying things, public policy evaluation in Latin America has followed a trajectory more or less parallel to that of Europe, and in which we can distinguish the following phases (Neirotti, 2012):

- An initial phase, which runs through to the late 1970s, associated with early development of the welfare state and characterised by development-oriented ("desarrollista") policies, designed under centralised planning systems. This phase is cut short in some countries with the arrival of authoritarian regimes (Chile and Uruguay in 1973, Argentina in 1976), although in other cases it coincides with such regimes (Brazil and Paraguay).

Although public policy evaluation as such still did not appear on the scene, it does start to emerge in the form of ex-ante analysis of the economic, financial and social viability of investment projects, with the support of the Organization of American States (OAS), the Inter-American Development Bank (IDB) and the World Bank (Feinstein, 2012).

This phase generated new sources of information necessary for the management and monitoring of development plans (population censuses, economic statistics, educational statistics, etc.), which were fundamental to ensuring the evaluability of policies.

- A second phase beginning in the 1980s and associated with the wave of New Public Management reforms, during which several countries in the region embark upon processes of transition to democracy (Chile in 1989, Argentina in 1983, Uruguay in 1984-89, Brazil in 1985-89, Paraguay since 1992, etc.). The foreign debt crisis in the 1980s accentuated and increased efforts made in the previous decade in relation to *ex-ante* evaluation, with monitoring and evaluation units being set up all across the region.

During this phase, major structural and sectoral adjustment programmes were undertaken. Evaluation as such continued play a minor role, whilst significant weight was placed on ex-ante analysis of projects. In addition, some mechanisms were created, such as the National System for Evaluation of Public Sector Management Results (SINERGIA) in Colombia and later the National Evaluation System (SINE) in Costa Rica, both of which were launched with the support of the Inter-American Development Bank (IDB). This phase also saw the creation of evaluation mechanisms

associated with the multi-year action plans of the Brazilian federal government and municipalities (Neirotti, 2012).

In Chile, the Ministry of Finance introduced performance indicators in 1993, followed by evaluations of government programmes based on reviews of the logical framework in 1996. In the late 1990s, this type of mechanism spread to other countries in the region (Feinstein, 2012). In all cases, the emphasis was placed on analysis of the results and impacts of public interventions within the framework of state modernisation processes, although they seemed to follow a logic related less to evaluation and more to the performance measurement and assessment of the satisfaction of users of public services associated with New Public Management.

Nonetheless, during this phase there was also increasing professionalisation of evaluation (Neirotti, 2012) and capacity-building actions were pushed by both the different governments in the region and international organisations, although more so by the latter and linked to development cooperation programmes at the state, regional or local level, as well as in sectors such as education, healthcare systems or social services. As we saw in different European countries, in this period there was also a rise in benchmarking practices in Latin America between municipalities, regions and service provider centres.

- A third phase, which began at the end of the last century and continues to present day, characterised by a 'return to the state' (which resumes its role of social and economic promotion although delegating some functions to the private sector), accompanied by new demands for institutional and democratic regeneration from an increasingly well organised and mobilised civil society. It is in this context that we see the emergence of evaluation approaches that place an emphasis on the participation of different stakeholders and incorporating different perspectives—not only those referring to the different levels of government involved in the evaluated interventions but also those of different social agents, ethnic groups, etc.—whilst also serving to raise awareness and provide training. At the same time, there is a more widespread adoption of *mixed methods* approaches centring on evaluating impacts or outcomes, but also capable of opening up the 'black box' of implementation to seek explanations for the levels of effectiveness achieved.

Generally, in Latin America there is currently a wide range of evaluation systems, often articulated around certain intervention sectors. These systems have usually been generated as a result of major programmes co-funded by donors in these sectors, and over time they have gradually become established within their respective institutional contexts through, for example, the creation of evaluation units in the government departments responsible for managing such programmes. These structures have in turn driven demand for evaluation and professionalisation in this area in the different countries, normally encompassing various different disciplines (sociology, anthropology, economics, psychology, healthcare, etc.). This professionalisation process has included the creation of several masters-level programmes, as well as evaluation societies of transnational dimensions, such as the Monitoring, Evaluation

and Systematisation Network for Latin America and the Caribbean (ReLAC), the United Nations Evaluation Network for Latin America and the Caribbean (EVALUN LAC), the Monitoring and Evaluation Network of Latin American and the Caribbean (REDLACME) and the Regional Platform for Evaluation Capacity Building in Latin America and the Caribbean (PREVAL).

1. Evaluation in Chile

Chile has a public policy monitoring and evaluation system which is often described as exemplary (May et al., 2006). This system has been developed since 1990⁵² within the context of a unitary state with a presidential political system, where legislative initiative in matters of finance and budgets rests exclusively with the President. This means that the executive branch is responsible for the administration and management of public finances, with the respective powers concentrated within the Ministry of Finance.

In this context, Chile has developed a complex monitoring and evaluation system, linked to management of the budget and imposed in a top-down manner from the Ministry of Finance and, more specifically, from the Management Control Division of the Budget Department (DIPRES). The Comptroller General of the Republic (equivalent to the European courts of audit) has not played a role comparable to that seen in some European countries due to the fact that, in Chile, this body does not have the power to audit performance.

Several factors have contributed to the design, implementation and gradual consolidation of the system, including the good health of the Chilean economy and its public finances in the early 1990s, which served to stimulate interest in achieving more effective and transparent application of public resources (Guzmán et al., 2014). Likewise, during this period a series of reforms were introduced in the management of public finances, including a review of the budget formulation and management process. There were also other initiatives aimed at modernisation of public sector management, such as the introduction of strategic planning, the development of information and communication technologies, typical NPM tools such as service charters, and, in general, improved information and attention to users. Furthermore, this period also sees regulation of Public Sector Senior Management, promotion of training for civil servants, and the introduction of performance measurement systems linked to salary incentives.

The monitoring and evaluation system currently covers all public bodies of the government and the central state administration included in the Budget Law and

52. Although the origins of evaluation in Chile date to the 1970s, with the incorporation of ex-ante evaluations in public sector investment projects, within the framework of what today is known as the National Investments System.

includes various monitoring, evaluation and management control mechanisms that have been gradually deployed and renewed in four main phases since the 1990s to present day (Arenas and Berner, 2010; Guzmán, Irarrázabal and de los Ríos, 2014). We will now go on to look at this evolution in more detail.

First phase: 1993 - 1999

In 1993, the Ministry of Finance began implementing monitoring mechanisms that allowed information to be generated on the performance of public sector programmes and organisations, taking as references the initiatives introduced in the United Kingdom under the governments of Margaret Thatcher as described in earlier sections of this report. The aim was to introduce practices that would improve the quality of spending and advance in matters of transparency and accountability (Arenas and Berner, 2010).

To this end, a strategic planning process was implemented in five public services (in the areas of education, agriculture, social services, treasury, and finance) along with the necessary IT systems to enable their management and to measure performance indicators. In 1994, this initiative was extended to 26 public sector institutions in Chile, setting in motion a system of performance indicators and management goals that would be incorporated into the process for formulating the Budget Bill of 1995. The results of these performance indicators were reported on a six-monthly basis by the executive branch to the National Congress. In 1997, the initiative was extended to some 67 services, representing almost 80% of all institutions eligible for inclusion in the system. The number of indicators practically tripled to 291 (Arenas and Berner, 2010).

During this time, the role of the National Congress in the debate of the Budget Bill gained significance, as it began to request information regarding the results of public sector institutions from the executive branch, through the Finance Commission. In 1996 the first Memorandum of Understanding was signed between Congress and the executive branch regarding the processing of the Budget Bill of 1997. This document set out a series of commitments aimed at increasing evaluation and transparency in public sector management. Among the initiatives arising from this memorandum was the commitment to evaluate government programmes on social issues, productive development and institutional development, as well as the obligation for public sector institutions to submit annual financial statements and management reports.

In turn, in 1997 this led to the addition of two new tools to the monitoring and evaluation system:

- The Government Programme Evaluation Programme (EPG), which evaluated some 20 public sector programmes during its pilot phase.

- Comprehensive Management Reports (BGIs), which contained the strategy statements of each institution, the results of certain performance indicators, and, predominantly, budget execution information.

It is important to underline the fact that, during this stage, the emphasis was placed on the design and gradual implantation of these tools, rather than the integration of decision-making information. This would gradually start to happen in a subsequent phase in the 2000s (Guzmán et al., 2014).

Second phase: Implementation of the Evaluation and Management Control System (2000-2005)

In this stage, the implementation of the tools seen in the previous stage is extended throughout the Chilean administration, and the tools themselves undergo some enhancements. In particular, these enhancements were intended to improve the strategic definitions contained in the Comprehensive Management Reports and the performance indicators, as part of the package of information submitted to Congress together with the Budget Bill. In addition, new lines of evaluation were created, with procedures introduced to ensure the dissemination of tools and their results, and to ensure their use in decision-making processes. More specifically:

- **Strategic definitions** must contain information on the mission, strategic objectives, strategic products (goods and/or services) and customers, users or beneficiaries of public agencies. The objective of including them in the budget process was to provide information on why and how resources would be used, thus contributing to a results-oriented budget discussion. Since 2000, public institutions have been required to report their strategic definitions on an annual basis as part of the yearly draft budget formulation process.
- These strategic definitions must be accompanied by **performance indicators** intended to measure the degree of progress of each institution towards the most significant objectives set. The performance indicators must serve as a way of quantifying and measuring results in relation to the goods and services provided by each institution to their respective beneficiaries, customers or users, monitoring the entire value chain: processes, products, intermediate outcomes and final outcomes. Furthermore, they must cover the different performance dimensions: efficiency, effectiveness, economy and quality. Moreover, the indicators must affect different areas of control, such as processes, products and outcomes, distinguishing between intermediate and final outcomes. However, in this stage, most of the indicators were centred on levels of production of goods and services, as well as on aspects of quality and especially user satisfaction levels (Mat et al., 2006). In contrast, impact indicators were not used in the system during this phase, given the greater methodological complexity required for their definition and measurement, and given the fact that

this type of indicator should be reserved for programmes with a certain level of maturity before they are measured.

In addition, performance indicators collected information over time, and therefore comparisons were mainly done with respect to past performance within the same organisation. During this phase, the comparison of performance indicators from different public programmes was somewhat problematic given that they all had different objectives, making it difficult to find consistent indicators (May et al., 2006).

The different public organisations themselves were required to prepare their respective BGIs directly, including the strategic definitions and performance indicators. The Budget Department of the Ministry of Finance had the role of coordinating the process and providing technical assistance, as well as of subsequently presenting the system of indicators to Congress for accountability purposes and using it in budget management and decision-making processes.

However, during this phase these instruments were seen by public sector organisations more as external control mechanisms than internal management tools, and they were therefore not fully 'internalised' by these organisations at the time. Likewise, the availability and quality of the information needed as inputs for the the different elements of the system showed not inconsiderable room for improvement (May et al., 2006).

- As regards **evaluations**, on the other hand, the Government Programme Evaluations (EPG) system, launched in 1997 and deployed during this phase, were centred on analyses based the matrix from the logical framework for the intervention in order to afterwards analyse their justification, design, organisation and management, as well as the level of performance (defined under the same terms used for the system of indicators) and their sustainability. Subsequently, in 2001, impact assessments would also be included, followed in 2002 by institutional evaluations, which evaluated a ministry in its entirety, including the set of goods and services it provides, as well as comprehensive spending evaluations.

At this stage, impact assessments would be carried out retrospectively and normally involve the use of quasi-experimental methods. They also had to fulfil a series of requirements:

- Firstly, they needed to be independent. As such, the institution with responsibility for execution of the programme could not conduct the assessments itself, and instead they had to be carried out by expert panels or consultancy firms outside the public sector and contracted and administered by the Ministry of Finance⁵³. The institutions participated throughout the evaluation process, but responsibility

53. The EPGs were conducted by teams of three professionals: an expert on evaluation, an expert on the intervention sector associated with the programme under evaluation, and an expert in public sector management. Impact assessments were commissioned from university centres or consultancy firms. Finally, the EPNs, which we will see in the next stage, were normally commissioned from universities, with the support of panels of international experts created on an ad hoc basis for each evaluation.

for the evaluation was with the Ministry of Finance as the client of the external consultants or consultancy firms.

- Secondly, the evaluation programme had to be transparent, in other words all information resulting from the evaluations would be of a public nature. Once the evaluations were concluded, they would be submitted to Congress together with the response of each of the evaluated institutions and an executive summary. This same information would also be published on the organisation's website.
- The evaluations had to meet certain technical requirements, and therefore a set of supporting documents, methodologies and procedures were developed in collaboration with experts such as evaluators and/or external advisers. The process for selecting these evaluators involved a public tender process and the setting of strict technical quality standards.

Between 1997 and 2004, some 158 EPGs were conducted, with an average cost equivalent to US\$ 11,000 and taking from four to six months to complete. The more exhaustive and rigorous impact assessments had an average cost of US\$ 88,000 and took eight months to complete. Up until 2004, 14 such evaluations were conducted, in other words around four per year. An analysis of the evaluation system developed during this phase found a certain disparity in the quality of both types of evaluations and pointed towards strict budgets and tight deadlines as possible causes. In this sense, some authors suggest the need to revise this model:

"Would the Ministry of Finance of Chile ever conduct as detailed and costly an evaluation as that of [Mexico's] Progres a [programme]? The danger for Chile is that its emphasis on keeping the cost of evaluations to a minimum perhaps leads it to under-invest in individual evaluations, even when the programmes being evaluated are costly government programmes with great political weight". (May et al., 2006, page 22).

As a general assessment of this phase (May et al. 2006):

- One feature of the monitoring and evaluation system that is seen as very positive is its gradual approach, which allows institutions to progressively adapt to new instruments.
- The creation of specific Administrative Control areas within public agencies is also seen as positive. These areas are given responsibility for internally coordinating the evaluation monitoring process and ensuring that the necessary information is available, etc. However, the fact that the different organisations do not have a specific budget for monitoring and evaluation activities is criticised.
- Also seen as a limitation is the tendency of institutions to formulate very basic targets and indicators, as they are close to being achieved already, particularly in cases where performance evaluation is linked to financial incentives for the institution and its employees. This is a common problem associated with performance indicator systems (Van Thiel and Leeuw, 2002).

- In addition, both the evaluations and the system of performance indicators refer to programmes defined for individual institutions, which does not fit well with the new generation of broad, cross-cutting social policies or with intersectoral interventions in which a policy involves several different institutions.
- Furthermore, criticisms are the shortcomings associated with a lack of buy-in to the system and the low level of institutional learning typical of systems implemented in a top-down manner, as well as the Ministry of Finance's excessive focus on the budget instead of results. That said, there are formal mechanisms for feedback and following up the recommendations included in evaluations, oriented towards facilitating learning and promoting the use of such recommendations for improving programmes. In fact, there are some fairly precise data available on the degree to which this information is used⁵⁴.
- Finally, the assessment also highlighted the still low level of coverage in proportion to public spending, as well as a lack of transparency associated with the dissemination of information in formats suitable only for expert audiences and not for the public in general.

Third phase: Consolidation and Strengthening of the Evaluation and Management Control System (2006-2009).

This phase saw the strengthening of the Evaluation and Management Control System, with the inclusion of corrective measures to address the shortcomings detected in the previous phase, as well as the incorporation of new instruments (Arenas and Berner, 2010). Specifically:

- The coverage of evaluations was broadened, with evaluation extended to 155 programmes and institutions between 2006 and 2010, representing half of all budgeted programmes.
- A new evaluation initiative was created in the 2009, known as New Programme Evaluation (EPN). This initiative is particularly interesting since it involves prospective evaluations of the impacts of new programmes, based on rigorous impact assessment methods, preferably using experimental designs. The prospective evaluation approach ensures the availability of the data needed for this type of evaluation, something which is, in most cases, impossible to achieve in retrospective evaluations.
- Furthermore, from 2007 onwards, technical assistance services were provided to public institutions in relation to the different instruments of the evaluation and monitoring system.

54. Between 2000 and 2004, it was determined that 25% of programmes needed minor adjustments, e.g. changes to the internal management, monitoring or other changes. In addition, 10% of the programmes evaluated were dissolved, whilst 21% of programmes underwent significant redesign of components or internal management processes. The rest involved changes in the programme design and internal management processes (39%), and the institutional relocation of the programme (5%). (Arenas and Berner, 2010).

- The design of and monitoring systems for new programmes and expansions to programmes presented in the budget have been reviewed since 2008.

In addition:

- A Quality Framework Programme was created and incorporated within the Management Improvement Programmes (PNG) in 2009, requiring public institutions to obtain ISO certification for their processes for the provision of goods and services.
- The Public Sector Management Modernisation Fund was set up in 2008 with the aim of spurring identification and resolution of weaknesses or problems in the management of public policies, programmes, instruments and institutions.

In general, according to the analysis conducted by the Budget Department of the Ministry of Finance (Arenas and Berner, 2010), the application of the system in the period from 2001 to 2010 showed that the coverage of the instruments had increased and that there was a higher level of expectation that institutions would fulfil their commitments⁵⁵. In this context, a system was created whereby, once the annual budget had been approved, DIPRES would work with each evaluated organisation on the implementation of the recommended improvements. These recommendations would be set out in formal agreements or Institutional Commitments (Guzmán et al., 2014).

Fourth phase: 2010-2014

Since 2010, the system of performance indicators has been simplified in terms of the number of indicators, and evaluations have placed the accent on measuring impacts, with less of an emphasis on analysing implementation. Attempts were also been made to improve the use of the information produced by the budget management system. In addition, new agents appeared on the scene: the Ministry of Social Development (MDS), created in 2011, and the Ministry General Secretariat of the Presidency (SEGPRES), with duties for coordinating and monitoring government priorities. The

55. In 2000, 72 public institutions adopted 275 indicators, whilst the level of attainment of only 59% of these indicators was evaluated; however, during formulation of the 2008 Budget, 142 public institutions adopted 1,443 indicators, 100% of which were evaluated. The quality of the indicators also changed: from 1,684 indicators adopted in 2004, 70% were associated with products or outcomes; whereas during the preparation of the 2010 Budget Bill, 20% of the indicators measured intermediate or final outcomes and 71% measured outcomes within the scope of the product. As regards evaluation, in the 1997-2010 period, some 255 programmes were evaluated through the EPG initiative, 85 programmes through impact assessments and 44 institutions through comprehensive spending evaluations. Furthermore, nine evaluations of new programme (EPN) were started. In terms of resources, between 2000 and 2005 the equivalent of 12% of the budget for that final year was evaluated. If we consider the evaluations undertaken between 2006 and 2010, the equivalent percentage is 34% of the budget for 2010. Finally, the tender processes of the Public Sector Management Modernisation Fund financed the preparation of 35 proposals to resolve weaknesses or management problems in public policies, programmes, instruments and institutions between 2008 and 2009. Moreover, the human resources employed in promoting and coordinating the monitoring and evaluation system within the Ministry of Finance increased significantly during this period: in 2000 the Management Control Department employed 11 professionals, a figure which later grew to 16 by the end of 2005 and 27 in 2010.

duties and resources of the Budget Department (DIPRES) have remained largely unchanged, except with respect to the monitoring and evaluation duties taken on by the MDS, which has taken over responsibility for the ex ante evaluation and monitoring of social programmes. The MDS has also set up a new mechanism: the Integrated Social Programmes Bank (BIPS)⁵⁶, which collects exhaustive information on social programmes, accessible to the public, including ex ante and retrospective evaluations.

We can therefore conclude that, in little over 20 years, Chile has managed to implement a monitoring and evaluation system with broad coverage in the central state administration. Among other things, the system includes various public policy evaluation initiatives, including rigorous impact assessments (both retrospective and prospective).

This has been achieved in a favourable context of sustained economic growth and political and social stability. And this was possible thanks to a combination of a strong political resolve with respect to the development of evaluation (which has continued throughout the different governments) and the existence of sufficient technical capacity and resources in the administration, particularly in the Ministry of Finance. In the initial stages, the highly centralised system was undoubtedly a contributing factor in extension of the model, although this centralised approach has also been successfully combined with incentives and technical training, particularly from the mid 2000s onwards, which have facilitated internalisation of the model.

The evaluation system in Chile

General assessment

Over the last three decades, Chile has deployed a sophisticated monitoring and evaluation system with broad coverage. The first steps date to the early 90s and follow the model of the administrative reforms implemented by the British government in the 80s, with an emphasis on performance measurement and a link to the budget process. The process is highly centralised in the Ministry of Finance and its Budget Department (DIPRES), and follows a top-down regulatory approach, particularly in the early stages. Congress takes advantage, almost from the very beginning and increasingly, of the opportunities provided by the new system for improving its monitoring and oversight over the executive branch. In 1997, the Government Evaluations Programme (PEG) is added to the system, with a 'desktop' evaluation approach which follows the analytical model of the logical framework. Later, in 2001, we saw the introduction of retrospective impact assessments commissioned from external experts, which would gradually extend their coverage over the course of the decade. In 2009 came the addition of New Programme Evaluations, which are prospective impact assessments that begin in the early stages of programmes, employing experimental methods where possible.

56. <http://www.programassociales.cl/>

The high level of centralisation of the system and the 'top-down' approach to its implementation initially caused problems such as a lack of buy-in and poor quality information from public institutions. The creation of specific units with responsibility for the system in each public institution, together with the technical assistance and training initiatives introduced in the late 2000s, helped to ensure better internalisation of the system. Moreover, the system of indicators has been gradually perfected, with the number of indicators reduced and an increased focus on measuring results and impacts. The evaluation approaches and methods have also been refined, with priority given to rigorous impact assessments. And there have been improvements in the use of the information produced by evaluations for the improvement of programmes, through the introduction of formal commitments and systematic monitoring. More recently we have seen the introduction of new agents: the Ministry of Social Development, with respect to the evaluation of social programmes, and the Ministry General Secretariat of the Presidency, with duties for the coordination and monitoring of government priorities.

2. Evaluation in Mexico

Systematisation of the practice of public policy evaluation in Mexico began in the late 1990s, coinciding with the start of a process of political change that enabled the country to make significant advances in democratic quality and carry out various administrative reform processes.

Mexico is a presidentialist federal republic comprised of 31 states plus the Federal District. Executive power is exercised by the President of the Nation, who is elected by direct universal suffrage every six years, with no possibility of re-election. Legislative power lies with the Congress of the Union, which consists of two chambers: the Chamber of Deputies and the Senate. Senators are elected every six years and deputies are elected every three years, with no possibility of immediate re-election in either case. The political organisation of the states reproduces this same system, with governors elected directly every six years and legislative elections every three. Despite the federal structure, the decision-making process is highly centralised both at federal level and at state level. Individual leadership is very important, and the fairly fluid mobility of senior officials leads to lack of continuity in institutional policy.

In the 1997 legislative elections, for the first time in the modern history of Mexico, the Institutional Revolutionary Party (PRI) did not win a majority. Later, in the 2000 presidential elections, political change was consolidated with the election of Vicente Fox, the National Action Party candidate (PAN). The PAN would remain in power after the 2006 presidential elections, which saw the election of President Felipe Calderón.

In the years leading up to this political and institutional transformation, the information available on social programmes mainly referred to the execution of the respective

budgets. There was no pressure on the executive branch from other state powers or any public demand for information on the results of public interventions. Between the 1970s and the early 90s there were several attempts to create evaluation schemes for federal programmes, generally associated with the Secretariat of Programming and Budget. However, these mechanisms did not progress due to a lack of consensus and, above all, due to the aforementioned lack of sufficient political and public pressure.

First phase: towards a results-based monitoring and evaluation system (1999-2006)

The mid 1990s saw promotion of a programme-based budgeting structure that continues to this day and resulted, among other things, in the Chamber of Deputies becoming more involved in monitoring and control of the actions of the federal government. Furthermore, in the late 1990s we saw the emergence of ad hoc evaluations of social programmes (e.g. the Social Milk Supply, Rural Food Supply, and Employment Training programmes) which generated partial information on the situation of certain elements of these programmes.

However, the milestone that revolutionised evaluation practice in Mexico was the set of quantitative and qualitative evaluations conducted for the conditional cash transfer programme known as *Programa de Educación, Salud y Alimentación* or PROGRESA. In 1997, the Mexican government launched a trial of the programme, something which was seen as very innovative back then: mothers would be paid a monthly stipend as long as their young children were up to date with all their vaccinations and preventive medical care and their older children attended school. Due to a lack of resources and certain logistical complications, the policy was unable to reach all areas of the country at once. The group of towns in which the policy was first implemented was selected randomly, which allowed the evaluation to follow an experimental design. The level of rigour in the evaluation, its independence and the valuable information that it kept the programme clear of political manipulation, and PROGRESA survived the sexennial change of government of 2000. The example of PROGRESA—later renamed OPORTUNIDADES and today known as PROSPERA—promoted the expansion of public policy evaluation in both Mexico and a good part of the rest of the American continent.

Although perhaps the most significant institutional change for progressing towards a culture of evaluation in Mexico was the law approved by the Congress of the Union in 1999 requiring the Mexican government to present annual external evaluations of all federal programmes subject to operating rules, in other words all subsidy programmes. This sudden change from practically no evaluation to the evaluation of hundreds of programmes every year, despite resulting in a slow process of learning and gradual implementation, did lead to many of the federal government's employees quickly becoming familiar with the practice of evaluation. Thus, in 2001 nearly 100 programmes

subject to operating rules were obliged to undergo external evaluations conducted by academic and research institutions with expertise in different fields, and these evaluations were presented each year to Congress as feedback for budget decisions.

Another noteworthy aspect in that period at the end of the 1990s, driven by the opposition-party majority in Congress, was the reform of the institutional framework for the national supreme audit function. This reform included amendment of various articles of the 1999 Constitution and enactment of the Federal Supreme Audit Act in December 2000. This new legal framework created the Federal Supreme Audit Office (ASF), reporting to the legislature and replacing the Audit Board of the Exchequer. Among its many duties, the ASF was responsible for conducting so-called performance audits aimed at assessing the degree of attainment of targets and objectives in government programmes, based on criteria of effectiveness, efficiency and economy, as well as assessing their social and economic impacts and the benefits generated for the public⁵⁷.

Almost at the same time, in 2001, an internal audit body within the executive branch known as the Secretariat of Comptrollership and Administrative Development (SECODAM) changed its role and name to become known as the Secretariat of Public Administration (SFP). The SFP continued to exercise its audit and control functions over the different federal institutions and departments⁵⁸, but it also focussed its attention on promoting various administrative reforms. It placed a particular emphasis on the areas of transparency, regulatory improvement, administrative simplification, e-government and professionalisation of the civil service (especially through the creation of the Professional Career Civil Service in 2003).

57. The ASF website (www.asf.gob.mx) contains only the so-called 'Public account reports' and other reports associated with audit activities for each year. Nevertheless, the information and reports viewed on the website of the ASF are synoptic in nature and have not allowed us to determine the actual scope of these actions.

58. The SFP has various different units with evaluation, audit and governmental control functions:

- A Government Performance and Management Evaluation Unit, the duties of which include the following (art. 24): (a) Establish, organise and coordinate the "governmental control and evaluation system". (b) Carry out the actions that, within the remit of the SFP, "are necessary for the correct operation of the performance evaluation system set out in the Federal Budget and Fiscal Responsibility Act, in order to ascertain the results of the application of federal public resources, the social impact of programmes and projects executed with public resources and identify the efficiency, economy, effectiveness and quality of the departments, agencies and the Office of the Attorney General". (c) Propose "general regulations, guidelines and policies to govern the operation of the governmental control and evaluation system". (d) Validate the management indicators which, in terms of the Federal Budget and Fiscal Responsibility Act, serve as the basis for evaluating the degree of attainment of the objectives and targets of the programmes and projects executed by the departments, agencies and the Office of the Attorney General, and the performance of these institutions in relation to the execution of said programmes and projects. (e) Establish evaluation criteria and evaluation methodologies or models that enable the operation of the governmental control and evaluation system. (f) Administer the information generated by evaluations, analyse the same and formulate proposals.
- A Public Sector Management Control Unit, the responsibilities of which include the following (art. 25): "establish, organise and coordinate the governmental control and evaluation system in collaboration with the Government Performance and Management Evaluation Unit".
- A Governmental Audit Unit responsible for an annual audit programme and conducting inspections of the public agencies of the federal government.

Also in 2001, but this time within the area of social policy, the Secretariat of Social Development (SEDESOL) set up a specialised agency comprised of a team of professionals with extensive technical backgrounds in evaluation, making the evaluation function a formal and permanent feature in this area of public policy.

Another step forward was taken in 2004 with the approval of the General Law for Social Development (LGSA), with a practically unanimous vote in favour by all political parties. The law sought to institutionalise various practices in social policies, including evaluation. In August 2005, SEDESOL published in the Official Journal of the Federation a decree creating the National Council for the Evaluation of Social Development Policy (CONEVAL) as a decentralised public agency of the Federal Administration with technical and managerial autonomy. Moreover, this declaration set out the objectives of CONEVAL as:

"To regulate and coordinate the evaluation of the National Policy for Social Development and the policies, programmes and actions executed by public departments, without prejudice to the responsibilities for evaluation and control held by the Secretariat of Finance and Public Credit and the Secretariat of Public Administration, and to establish guidelines and criteria for the definition, identification and measurement of poverty, whilst guaranteeing the transparency, objectivity and technical rigour of such activities." (Art. 3)

The law created CONEVAL as an institution with technical and managerial autonomy with the mission of evaluating social development policy, its programmes, funds and actions, and of measuring poverty at national, state and municipal level. The institution is responsible for evaluating and coordinating the evaluation of social programmes and policies in all the country's federal government secretariats.

Pursuant to the General Law for Social Development (2004), the CONEVAL's governing board is chaired by the Secretary of Social Development and consists of six academic researchers elected through a process including public tender and the vote of the 32 governments of the federal entities, the Chamber of Deputies, the Senate, and representatives of the municipalities and the Federal Government. The governing board also includes a representative of SHCP and two members of the Supervisory Body. The weight the six academic researchers have in governance of the institution gives it technical independence and strengthens the credibility of its actions.

The board also achieves a reasonable level of financial independence thanks to a budget which is seen to be sufficient to promote the necessary evaluations of social programmes, at least in those early years (González de Licona, 2010).

To summarise this stage we can say that, between 2000 and 2006, around 500 external evaluations were conducted, aimed at Congress, the SHCP, the SFP and the managers of the evaluated programmes. These evaluations were generally made public.

Nonetheless, their utilisation for improving public policy was, according to some of the those involved in the process, practically non-existent (González de Licona, 2010).

Second phase: consolidation and expansion of the monitoring and evaluation system (2007 - present)

The Budget and Fiscal Responsibility Act (LPRH), enacted in 2006 and reformed in 2007, created the Performance Evaluation System (SED) for the purpose of evaluating the performance of public programmes and projects in order to contribute to a results-based budget (RBB). In practice, this performance evaluation system, developed by the Secretariat of Finance and Public Credit (SHCP), was a set of indicators, evaluations and methodologies created to provide objective information to assist in taking budgetary decisions. The SED was applicable at the federal level and therefore affected all public programmes funded at this level, including social programmes. But the law also stipulates that the states and municipalities should advance in the area of monitoring and evaluation.

The Federation Expenditure Budget Decree published by Congress in January 2007 stipulated that the coordination of the evaluation of federal programmes would fall under the remit of CONEVAL, the SHCP and the SFP. The decree also stated that, to this end, General Guidelines for Federal Programme Evaluation would be published. These guidelines were later published in March 2007, with the explicit objective of regulating the system for monitoring federal programmes for the purpose of standardising the types of activities to be carried out and helping to consolidate the SED. Hence, these reforms were aimed at establishing a link between the planning, the budget and execution of federal programmes through the use of monitoring and evaluation tools.

In addition, the 2007-2012 National Development Plan, a plan required by the Planning Act once every six years and coordinated by the SHCP and the Office of the President, included two innovative elements:

- Clear and quantifiable targets and, for most of these, strategic outcome indicators to reflect improvements to the well-being of the population.
- The highest possible level of correlation was sought between the National Development Plan and the various sectoral programmes, for which quantifiable targets and strategic outcome indicators were also devised. This same system of objectives and strategic indicators was also established for each of the federal government's secretariats and departments. These targets and indicators were set through a process of close collaboration between the secretariats and the Office of the President of the Republic.

The 2013-2018 National Development Plan, for its part, will maintain and strengthen these monitoring and evaluation elements, as well as elements for coordination and integration between different planning levels (CONEVAL, 2014).

As regards specific aspects of the post-2007 monitoring and evaluation system, the General Guidelines for Federal Programme Evaluation promoted the following:

- First of all, in order to improve the internal logic of programmes, better define their ultimate purpose and devise outcome indicators, all federal programmes were required to have a Logical Framework Matrix, which in this context was referred to as a Matrix of Indicators. This matrix would also serve to explicitly define the connection between the different programmes and the national objectives expressed in the National Development Plan.
- In addition, in order to achieve more coordinated and consistent evaluation of federal government policies and programmes, the Guidelines proposed the idea of an Annual Evaluation Programme that would include the most important evaluations to be conducted each year.
- Furthermore, instead of having just one type of evaluation, as had been required since 2000, the Guidelines introduced different instruments with the aim of making the overall evaluation mechanisms more flexible. Specifically:
 - Evaluation of Consistency and Outcomes: to systematically assess the design and overall performance of programmes based on a matrix of indicators.
 - Evaluation of indicators: to assess the pertinence and scope of programme indicators through fieldwork.
 - Evaluation of processes (or implementation): to assess the effectiveness and efficacy of the operational processes for improving management.
 - Impact assessment: to measure the degree of effectiveness reached, attributable to the programme, using control groups to separate out the contribution of other agents in the attainment of final outcomes.
 - Strategic evaluation: focussing on a programme or group of programmes centring around a strategy, policy or institution.
 - Specific evaluation: conducted by the institutions or departments themselves on a voluntary basis to assess a specific aspect of their programmes or their management.

In the previous phase, the evaluations requested by Congress were not consistent, and as such the information could not be used to draw comparisons between programmes. These Guidelines sought to ensure that evaluations contained consistent elements that would improve the quality control of evaluations and provide aspects that could be used for comparing programmes.

According to the 2014 Social Development Policy Evaluation Report (CONEVAL, 2014), between 2007 and June 2014 some 1,504 evaluations were conducted on social

development programmes, actions and policies, and quarterly indicator reports were regularly published on the CONEVAL website.

This phase also saw the introduction of mechanisms to promote use of the information generated by the monitoring and evaluation system to improve the evaluated programmes and policies. Thus, in 2008 CONEVAL, the SHCP and the SFP published the Guidelines for the Continuous Development of Public Sector Performance. This instrument sought to ensure that the findings contained in evaluations, as well as those generated by other mechanisms put in place by the SFP and the SHCP, would be used to create specific action plans to improve performance. This instrument contained not only commitments to be met by the direct managers of the programmes but also commitments for institutions and other agencies outside the executive branch within which these programmes operate (such as the SHCP and other departments), who would be required to work together to improve performance in the public policy in question.

From this moment forward, CONEVAL would monitor all modifications made to programmes based on evaluations. According to information issued by the governing board (CONEVAL, 2014), significant changes have been made to social programmes since 2010. The most recent data shows that in 2013-2014, 47 programmes underwent a substantial refocusing (50%), whereas 39 programmes corrected certain activities or aspects of their operation (41%). Between 2010 and 2014, only one programme was suspended as a result of an evaluation.

The change in evaluation culture also seems to have spread to federal entities. CONEVAL has a system for assessing the operation of monitoring and evaluation systems at state level and, although things have historically been slower and more unequal among federal entities, it shows that substantial progress has been made over the last four years (CONEVAL, 2013).

Main challenges in the process of institutionalising evaluation

The main challenges identified are as follows:

- The persistent weaknesses in planning, especially within the field of social policies, despite the improvements recorded:
- *"...the objectives of the social development policy are defined in a fragmented and scattershot manner; the basic rights to which an individual should have access are not adequately defined in current legislation, whether this be in the Constitution or in secondary laws".* (CONEVAL, 2014, page 101)

- The shortcomings that persist—despite the progress made—in the integration and coordination of the different agents that form part of the planning, monitoring and evaluation system, particularly within the Office of the President, the SHCP, the SFP and CONEVAL, all of with responsibilities in the promotion, supervision and definition of performance monitoring and evaluation matters.
- Moreover, it is important to ensure that information is generated in order to guarantee optimal evaluability conditions. In this respect, it is also important to strengthen the technical capabilities of professionals in charge of programmes and of planning and evaluation areas in federal departments; improve internal capacities for the analysis of basic programme information, not just budgetary information; continue with existing lines of action in order to improve administrative records as a way of supporting external monitoring and evaluation activities; and standardise said administrative records in close collaboration with the National Institute of Statistics and Geography by making use of statistics committees and generating information to be coordinated by the INEGI in the different public sector agencies and departments.
- Finally, there is a need to continue promoting the use of evaluations, not only to improve the different programmes and policies but also to improve general planning (in terms of reviewing objectives and monitoring indicators), as well as for the definition of budget programmes and for managing the budget (planning, designing, budgeting, and many other uses).

The evaluation system in Mexico

General assessment

The development of a public policy monitoring system in Mexico begins in the mid-1990s after the introduction of programme-based budgeting by the Secretariat of Finance and Public Credit (SHCP). In the later part of that decade, the success of the PROGRESA evaluation programme marks a turning point. In 1999, Congress approves a law which requires the federal government to present annual external evaluations of all subsidy programmes. Between 2000 and 2006, almost 500 external evaluations were conducted. Although only marginally used to improve the evaluated programmes, this phase helped to spread the evaluation culture in the public sector. In parallel, 2001 also saw the creation of the Secretariat of Public Administration (SFP), which promoted several administrative reforms that would help create a more favourable environment. Also in 2001 the Secretariat of Social Development (SEDESOL) started a process for the institutionalisation of evaluation in the field of social policy, a process that culminated in 2005 with the creation of CONEVAL. This institution is responsible for evaluating—and coordinating evaluation—of social programmes and policies in all of the country's federal government secretariats. The connection between the different components of monitoring and evaluation becomes a reality after passage of the Budget and Fiscal Responsibility Act in 2006, later reformed in 2007, which created the Performance Evaluation System (SED). The SED establishes that coordination of federal

programme monitoring and evaluation would fall under the remit of CONEVAL, the SHCP and the SFP. These coordination activities, and the practice of monitoring and evaluation, is governed by the General Guidelines for Federal Programme Evaluation published in March 2007. These Guidelines set out actions for improving policy design, establishing the use of a matrix from the logical framework as the method of reference for guaranteeing the internal logic of programmes. They also promoted an Annual Evaluation Programme (PAE) which included different evaluation instruments—to analyse consistency and outcomes, impact, processes, etc.—allowing the different public bodies margin to perform evaluations on their own. Also introduced in this phase were mechanisms to promote use of the information generated by the system, which is now regularly published on the CONEVAL website. Challenges identified as yet to be addressed include improvements to policy planning and design, better integration of the agents promoting the system, better technical evaluability conditions, and increasing the effective use of evaluation results in decision-making processes.

3. Evaluation in Colombia

The current model of institutionalisation of evaluation in Colombia began as a result of a process of political change which culminated in the 1991 constitutional reform. However, it is important to note some interesting background details.

The national planning model introduced in the 1950s started to show signs of fatigue in the 80s, when the urgent need for its reformulation became apparent after the increase of investment budgets and their negligible impact on the country's economic and social development. Given this concern, since 1990 management agreements have been introduced as mechanisms for validating results for the beneficiary entities of loans granted by multi-lateral bodies which were guaranteed by the government.

Furthermore, in the late 1980s, the National Planning Department (DNP), a body with ministerial rank created in 1958, launched two initiatives that were the forerunners of the current monitoring and evaluation system (DNP, 2010):

- The Project Evaluation and Monitoring System (SISEP), which fell under the remit of the DNP itself and was later integrated into the Public Investments Unit, today the Investments and Public Finances Division.
- The Project Bank, created by Law 38 of 1989, with the role of "evaluating all investment projects financed with central government resources, with an end to ensuring they fulfil basic criteria of financial profitability and technical viability".

In 1990-1991, the country underwent a process of political change which culminated with the reform of the 1886 Constitution. Among other aspects, the new constitution abolished the former restrictions on open competition between the different political

forces and, alongside the traditional electoral mechanisms, established participatory democracy mechanisms such as the open council or '*Cabildo Abierto*' (a public meeting of local government bodies), the legislative initiative or the option to revoke the mandate of representatives elected by the public.

However, the 1991 constitution also significantly broadened the constitutional rights of citizens, adding social and economic rights as well as so-called third-generation or collective rights. Other amendments that helped to create an environment more favourable to evaluation were the de-politicisation of the Office of the Comptroller General of the Republic and the decreased powers of the President in favour of the legislative branch (Alcántara, 2013).

The Constitution also established two guiding principles for national planning: the first being the participatory nature of the process for formulating development plans; the second being the principle of continuous evaluation of plans and public sector management, in terms of results. (Ocampo, 1996, cited by Villarreal, 2007).

First phase: 1994-2002. The National Development Plan Act and the launch of SINERGIA

Articles 343 and 344 of the 1991 Constitution of Colombia set out the framework for designing and organising systems to evaluate the management and results of public administration⁵⁹. Following this promulgation, evaluation is introduced as a public sector management instrument and is no longer a voluntary or isolated exercise conducted by some institutions (DNP, 2010).

Of particular note among the series of legal regulations that implemented and regulated the monitoring and evaluation process in Colombia, are Decree 2167 of 1992, Law 42 of 1993 and Law 152 of 1994.

Decree 2167 of 1992 restructured the National Planning Department, giving it the following responsibilities, among others:

- create the methodologies for preparation and evaluation of the programmes and sub-programmes of the National Development Plan;
- design and organise systems for evaluating the management and results of the administration, with regard to both policies and investment projects, and indicate

59. Article 343.- *"The national planning entity stipulated by the law will be responsible for the design and organisation of systems for evaluating the management and results of the public administration, with regard to both policies and investment projects, under the conditions it deems necessary".*

Article 344.- *"Departmental planning bodies shall evaluate the management and results of the development and investment plans and programmes of the departments and municipalities, and shall participate in the preparation of the budgets for the latter, under the terms stipulated by law. In any case, the national planning body may choose to conduct such an evaluation on any territorial entity".*

- the terms, conditions and the parties responsible for conducting the evaluation; and
- organise and implement a system for retrospective evaluation of the National Development Plan and for the plans of territorial entities.

For its part, Law 42 of 1993 makes reference to fiscal control and the way in which it should be exercised by the Office of the Comptroller General of the Republic. Furthermore, Law 87 of 1993 was passed in that same year, establishing regulations on internal control in state entities.

1994 saw ratification of the Organic Development Plan Act (Law 152/94), under the Presidency of Ernest Samper, for the first time assigning the status of law to this planning instrument which had been present in Colombia for over three decades. In accordance with the Constitution, the law makes the National Planning Department responsible for "designing and organising systems for evaluating the management and results of the administration, with regard to both policies and investment projects, and indicating the terms, conditions and parties responsible for conducting evaluation. Said systems shall take into account the attainment of targets, the coverage and quality of services, and the unit costs, and shall establish obligations and procedures for the provision of information by entities". (Art. 29).

Also in 1994, the National Council for Economic and Social Policy (Conpes)⁶⁰ approved Conpes Document 2688 on the evaluation of public sector results at the national level and issued a resolution in the same year creating the National Public Management Results Evaluation System (SINERGIA). SINERGIA is conceived as a national system incorporating various different agents (government agencies, agencies exercising control over the executive branch, representatives of civil society, the media, academia, etc.) in monitoring and evaluation activities.

Subsequently, in 1995, the Council for State Modernisation and the Special Evaluation Division (DEE), attached to the DNP, issued Conpes Document 2790 on results-oriented public sector management, which highlighted self-evaluation by agencies in charge of public programmes as one of the key elements of SINERGIA. In line with this strategy, a process is created whereby national public administration agencies are required to submit four-yearly indicative plans, which are intended to be used as strategic tools for setting objectives and for monitoring and evaluating the degree of progress achieved in relation to the targets set in the the National Development Plan⁶¹. In September

60. Body created in 1958 as the maximum authority on national planning, acting as a advisory body to government on all aspects associated with the country's economic and social development. With this mission, it coordinates and trains the bodies responsible for managing economic and social issues in the government, through researching and approving documents on the development of general policies which are subsequently presented in session. The DNP acts as an Executive Secretariat of Conpes.

61. Indicative plans had to contain the PND objectives associated with the agency in question, the priority objectives of the agency in order to reach them, the efficiency and effectiveness indicators (in terms of budgetary

1995, a trial project for implementation of indicative plans was set up in 12 agencies at national level (DNP, 2010).

As a result of the above, inter-agency working groups called 'Management Units' were created with the aim of promoting the cross-cutting nature of the system and agreeing on objectives for the different programmes of the National Development Plan, ensuring coordination of all the bodies involved and joint monitoring of levels of attainment. Furthermore, the involvement of the Ministry of Finance was also sought in order to ensure consistency with budgetary planning. However, although this strategy provided members of the participating units with the possibility of a more complete overview of their work and helped to resolve certain issues associated with inter-agency coordination, the initiative finally came to end as a result of its weak link to budget allocations and the difficulty of reaching agreements on the powers and responsibilities of each agency (DNP, 2010).

The first Conpes document regarding the monitoring of the National Development Plan, issued in 1997⁶², proposed a new mechanism, 'Efficiency Agreements', aimed at improving coordination between planning departments, as well as improving programme management and budget management. These agreements were intended to ensure a greater level of certainty with regard to available budgetary resources, setting clearer spending priorities and advancing towards attainment of the expected results.

However, and despite the fact that in 1997 all public agencies with responsibility for meeting the objectives set out in the National Development Plan (around 170) had already made an effort to define objectives and indicators within the framework of SINERGIA, the information available on both planning and monitoring was lacking in terms of quality and feasibility. In addition, problems associated with the links between the plans, the monitoring and evaluation mechanisms and budget management continued to exist (DNP, 2010).

In the late 1990s, the process begins to be affected by severe budget restrictions imposed as a result of the financial crisis that arrived at the end of the century. Added to these difficulties was the weight given to national security and defence on the political agenda in this period. Thus, between 1998 and 2000, the system reached a low point in terms of activity, political backing and institutional structure, with the implementation process coming to a standstill.

and financial resources), etc.

62. Conpes Document 2917. *The Social Leap*. National Development Plan Progress Report. 1997.

Second phase (2002 - present): system adjustment and consolidation

The breakthrough of the results-based management culture in the Colombian Administration arrived during the first term of office of President Uribe between 2002 and 2006. In addition to its priority objective of restoring security to the country, the Uribe Administration introduced an agenda of reforms that revolved around three pillars (DNP, 2010): a new culture of public management, deepened decentralisation, and strengthened public participation.

As regards results-based management, a process for the reform of SINERGIA was initiated, focused on resolving the weaknesses identified in the previous period in relation to the link between the system and budget management, as well as to the monitoring and evaluation conceptual framework, the quality and reliability of the information obtained and the frequency with which said information is produced, as a way of ensuring an optimal link with decision-making processes.

This being the case, within the framework of the Public Administration Reform Programme (PrAP) established based on the strategic objectives set out by Presidential Directive No. 10 of 2002, SINERGIA commenced a process of renewal that sought to strengthen the system around three major objectives: 1) contribute to improving the efficiency and impact of policies, programmes and institutions, 2) increase the efficiency and transparency of programming and allocation of resources, and 3) stimulate transparency in public sector management by actively incorporating citizen oversight.

More specifically, the system developed the following components:

a) The monitoring and oversight component

With this component, from 2002 onwards SINERGIA started to produce short term information on the results of the policies and programmes reflected in the National Development Plan through the use of a technological platform known as SIGOB (*Sistema de Gestión y Seguimiento a Metas de Gobierno* or System for Managing and Monitoring Government Targets).

The original design of SIGOB came from the United Nations Development Programme, within the framework of its regional project entitled 'Management for Governance'. It consisted of an online system that gathered data from different ministries and administrative bodies and which also lent transparency to the management process through the publication of results and targets. That system was in operation between 2003 and 2010, replaced in 2011 by a system called SISMEG (System for Monitoring Government Targets), coinciding with the entry into office of Juan Manuel Santos as president in succession to Álvaro Uribe after the 2010 elections.

The introduction of SIGOB sought the integration of information on targets and results of the main institutions of the executive branch over a four-year period. The specific objectives of SIGOB were as follows:

- Serve as a management information and control tool for institutions on behalf of the president.
- Provide the public and all agencies and branches of the government authority with simultaneous and transparent access to government results.
- Consolidate basic information for the purposes of the ex-post analysis of the main public policies, with the aim of strengthening subsequent years and future planning.
- Facilitate oversight of the management of agencies responsible for meeting the targets in the National Development Plan and government commitments.

In the mid 2000s, the system contained around 500 indicators associated with 320 National Development Plan objectives. For each indicator, the system recorded the objective, the strategies being implemented to meet the objective, baseline information, annual operational objectives, results based on these objectives, and the resources assigned and allocated in each case. The information was broken down by region and also by certain main cities. In addition, in order to strengthen accountability, in cases where the objectives were not met, the responsible manager was required to provide an explanation of the reasons why, and this information would be added to the system. All managers in charge of programmes were identified in the system, including details of the agency in which they worked and their contact information (Mackay, 2007).

Furthermore, as regards areas for improvement, although the system was widely used by the Office of the President of the Republic to monitor the government programme, the information itself had issues in terms of its quality, reliability and bias, as well as being inconsistent. Part of the problem stemmed from the fact that the information was entered directly into the system by the different public administration agencies themselves, with no system of audit or quality control.

As mentioned, SIGOB was replaced by SISMEG in 2011. This new tool was also a system for monitoring government planning, but with some differences with respect to its predecessor. Firstly, SISMEG is a more technologically advanced platform which, in addition to having a monitoring system, includes two products available for consultation by the public: the Economic Log and the International Comparison⁶³. The system issues regular reports summarising results at the national and territorial level, as well as an annual report for the Congress of the Republic As well as objectives associated with the monitoring and control of internal management and accountability,

63. <https://sinergia.dnp.gov.co/portaldnp/>

SISMEG also promotes a better link between planning and allocation of budget resources.

Throughout this entire period, the DNP has worked in conjunction with the Office of the President to implement actions to improve the quality and updating of the information contained in these monitoring systems, and to incentivise their use. These actions have included training for public bodies on monitoring issues, direct guidance, publication of manuals and guidelines, controls exercised over agencies during accountability and management activities, preparation of documents for the purpose of monitoring results, and starting to coordinate SISMEG with other information systems.

b) The evaluation component

From 2002 onwards, we start to see the development of the evaluation component of SINERGIA, with the aim of improving the effectiveness, design and execution of the policies and programmes included in the National Development Plan.

The evaluations to be conducted were included on evaluation agendas valid for a period of one year. Putting this evaluation agenda together is primarily a technical exercise carried out jointly by the technical divisions of the National Planning Department and the different government agencies, involving identification of the priority issues based on strategic objectives set out in the National Development Plan and the evaluation commitments undertaken in relation to programmes financed by multi-lateral bodies. This process is led by the Public Policy Monitoring and Evaluation Division (DEPP) of the DNP.

Unlike CONEVAL (Mexico), the DEPP does not enjoy technical, managerial and financial autonomy but, rather like DIPRES (Chile), it is an administrative body which, for all intents and purposes, forms part of the DNP structure. That said, the DEPP does not have the enforcement powers of the Chilean body, directly associated with budget management.

In 2007, with the objective of providing sustainability and continuity to this operating method, article 132 of Law 1151 of 2007 (the 2006-2010 National Development Plan Act) created the Intersectoral Evaluation and Results-based Management Committee (CIE) with the aim of defining the multi-year evaluation agenda based on public agency proposals. Furthermore, criteria for prioritising the programmes to be evaluated were defined, along with the methodologies to be used.

Once the agenda is set, the evaluations are normally designed by the DEPP or directly by the agencies responsible for the programmes. However, the percentage of evaluations conducted with the involvement of the DEPP with respect to the total

number carried out has seen a continuous increase, reaching almost 90% by 2010⁶⁴. The evaluations are normally conducted by academic experts or consultants, to whom work is awarded through a tender process. The internal manuals and guidelines on how to conduct evaluations stress the importance of involving the stakeholders in the evaluation process, from the design phase through to the implementation process and in dissemination and debate of the results.

The evaluation system distinguishes various types of evaluation, emphasising operational and institutional evaluations, on the one hand, focussing on operational and management aspects of programmes, as well as their institutional context; outcome evaluations, and impact assessments, the latter of which are more methodologically demanding and involve the construction of counterfactual scenarios. It also emphasises executive evaluations, which are quick reviews of programmes in operation to verify the quality of their design, how they are functioning, the degree of coverage achieved, etc. Between 2006 and 2011, 19 executive evaluations, 22 impact assessments, 7 institutional evaluations, 5 operational evaluations and 8 outcome evaluations were conducted (61 in total) (DINP, 2010).

The sectoral distribution of evaluations has gradually broadened over time: in 2004, 44% of evaluations referred to social protection programmes, with the rest referring to the environment, housing and territorial development sector. In contrast, in 2010 evaluations also included sectors such as agriculture, education, transport and infrastructures (DNP, 2010). In terms of the volume of evaluations, up until 2009 the evaluations agenda led by the DEPP included over 60 evaluations, with an investment of \$15 million. The number of evaluations completed in 2008 increased by 620% with respect to 2004. In recent years, the number of evaluations included on the agenda each year is around 30⁶⁵.

Since 2011, we have seen the introduction of a new strategy called the National Evaluation System (SISDEVAL). The objective is to methodologically and procedurally strengthen evaluations—through the adoption of an 'effective evaluations' component aimed at improving the technical capabilities of the public and private agents involved in the evaluation process—and to catalogue and publicly disclose the evaluations conducted at all levels of government—through the use of the 'evaluations radar'.

Also worthy of note is the creation, in 2009, of the Colombian Public Policy Monitoring and Evaluation Network⁶⁶, with the objective of establishing links between the agents involved in the design, development and evaluation of public policies in Colombia, especially at local level.

64. See DNP (2010). Results-based public sector management in Colombia. SINERGIA. Presentation given in Zacatecas, 27 April 2010. Available at <http://bit.ly/17nl7Ts>

65. See Evaluation Agenda 2013 DNP, DEPP, at <http://bit.ly/1B2Pz2h>, and for 2014 at <http://bit.ly/1MLZRbX>

66. <http://redcolme.ning.com/>

General assessment and main challenges yet to be addressed

The main challenges yet to be addressed are:

- Use of evaluation in the decision-making processes of public bodies with regard to the policies and programmes for which they are responsible, as well in the budget management process, still has significant room for improvement even though some progress has been made, including the standardisation of recommendations in the form of improvement action plans. That being said, progress still needs to be made on the monitoring of these actions.
- The direct link of the DEPP with the executive branch is sometimes seen in certain social spheres and by the media as a source of bias and of a general lack of independence in the monitoring and evaluation system.
- In some cases, the quality of the information fed into the system by the different public agencies is still poor. There are also difficulties in terms of the availability and quality of data in some sectors of public intervention.
- Finally, the funding of the evaluation agency still largely depends on multi-lateral donations, which threatens its medium and long-term economic sustainability.

The evaluation system in Colombia

General assessment

Promotion of evaluation in Colombia begins with the process of political change that culminates in the enactment of the 1991 Constitution, which explicitly assigns responsibility for evaluation to the body responsible for national planning—the National Planning Department. In 1994 the Organic Development Plan Act is passed, and the National Public Management Results Evaluation System (SINERGIA), which incorporated a monitoring component and an additional evaluation component, is launched. However, deployment of SINERGIA grinds to a halt in the late 1990s and does not resume until 2002, when the inauguration of President Uribe brings about a push for improving result-oriented public sector management. Unlike in Chile, the evaluation and monitoring system in Colombia was designed to have limited executive powers. However, the Public Policy Monitoring and Evaluation Division (DEPP) of the DNP has legitimised and consolidated its role as a leading evaluation body capable of advising and providing methodological support to different government agencies in their evaluation activities (defining terms of reference, commissioning and management of evaluations, methodological advice on impact assessments, etc.). The DEPP has followed a strategy of gradual approach, based on voluntary participation, raising awareness and gradually convincing government bodies of the value of evaluation for improving the design, operation and effectiveness of programmes. The level of dissemination of the results of the monitoring and evaluation system and use of the information derived from SINERGIA by the Office of the President and Congress seems high. However, there is

room for improvement in terms of its effective use in the improvement of programmes and allocation of budget resources, as well as the quality of the information fed into the system in certain cases. Finally, the largely external funding for the system presents a challenge for its sustainability in the medium-term.

Table 3. Overview of the evaluations systems described

Characteristic parameters of the evaluations systems	Germany	Belgium	Chile	Colombia	Spain	France	Italy	Mexico	Netherlands	United Kingdom	Sweden
Conceptual framework											
– Public policy evaluation practice differentiated from monitoring practice.	•	?	•	•	•	•	•	•	•	•	•
– Impact assessments of experimental or quasi-experimental design.	•		•	•	•	•	?	•	•	•	•
– Focus on control	•	•	•	•	•	•	•	•	•	•	•
– Focus on learning and improvement	•	•	•	•	•	•		•	•	•	•
Institutional framework - regulations											
– Explicit policy on public policy evaluation.			•	•		?		•	?		
– Reference to evaluation in cross-cutting regulations (budget, planning, etc.).	?	?	•	•	•	•	?	•	•	•	?
– Evaluation clauses in sectoral legislation.	•	•	?	?	•	•	?	•	?	?	?
Institutional framework - organisation											
– Monopoly			•								
– Centralised pluralism				•		•		•	•	•	
– Pluralism	•										•
– Fragmented concurrence		•			•		•				

Table 3. Overview of the evaluations systems described (cont.)

Characteristic parameters of the evaluations systems	Germany	Belgium	Chile	Colombia	Spain	France	Italy	Mexico	Netherlands	United Kingdom	Sweden
– Existence of differentiated evaluation units in public agencies			•	•	•	•	•	•	•	•	•
Evaluation practice - execution											
– Internal, centralised					•	•		•		•	
– Internal, decentralised		?			•	•	•		•	•	•
– External	•	?	•	•	•	•	•	•	•	•	•
Evaluation practice - quality											
– Application of common standards			•	•				•		•	
– Expert supervision	•		•	•		•		•		•	
– Data availability and quality	•	?	?	?		?	?	?	•	•	•
Evaluation practice - continuity											
– Evaluation agendas	?		•	•	•	?		•	•	?	?
– Regular execution of evaluations	•		•	•		•		•	•	•	•
– Specific budget	•	?	•	?	•	?	?	?	?	?	?
Public disclosure											
– Generalised dissemination	•	?	•	•	•	•	?	•	•	•	•
– Restricted dissemination		?					?				
– No public dissemination		?					?				
Effective use											
– For the improvement of programmes	•		•	•		•		•	•	•	•
– For parliamentary oversight of the executive branch	•		•	?		•		•	•	•	•
– For public oversight of the executive branch			?	?		?		?	?	•	•

Table 3. Overview of the evaluations systems described (cont.)

Characteristic parameters of the evaluations systems	Germany	Belgium	Chile	Colombia	Spain	France	Italy	Mexico	Netherlands	United Kingdom	Sweden
– Instrumental, as a tool for building consensus		•									
– Low level of effective use		•			•		•				
System evaluation											
– Existence of independent bodies that regularly assess system functioning.									•	•	•
– Internal monitoring by the agencies responsible for the system.			•	•				•	•		
– No statistical monitoring	•	•			•	•	•				
• Based on the documentation consulted, it appears that the parameter in question exists in the country*											
Based on the documentation consulted, it appears that the parameter in question does not exist in the country*											
? Based on the documentation consulted, it is impossible to establish whether or not the parameter in question exists in the country*											

(*) The information reflected in the table is based on an analysis of secondary sources conducted during the preparation of this report. Said sources do not allow us to capture in all cases, or with the same level of certainty and detail, information on the components of the evaluation system identified in the literature. A rigorous examination of the opinions reflected herein has been left for a later stage. Such an examination should be based on a critical review of the components of the evaluation systems identified in the literature in order to subsequently establish, in a reasoned manner, the variables and indicators of each system, as well as the information gathering methods that enable them to be correctly, consistently and systematically measured.

Source: Compiled by the author.

V. Conclusions and recommendations

The successful development of evaluation systems does not so much require a technical or institutional change but rather, and above all, a cultural change.

That **cultural change** must start from the grassroots, that is, by reasserting the values that should guide political action and orient the functioning of public administrations. Political elites should prioritise and encourage reflection and debate on what constitutes general interest and what public policies can best serve it. A difficult and divisive debate in which a consensus will rarely be reached and where democratic rules must determine which options will govern the community at any given time. But a debate that is infinitely more necessary than the purely tactical and short-term confrontations that alienate citizens from their political institutions.

This cultural change is not simple, since it goes against a political tradition in which excellence has less impact than failure, and where punishment of failure provokes risk aversion and therefore results in less opportunities for experimentation, learning and improvement:

“... within the political arena, the punishment for mistakes is severer than the reward for excellence.... Mistakes are more newsworthy and therefore get much more attention. This leads to slightly neurotic tensions in public organizations to avoid mistakes as much as possible. It is very well possible to have a successful career in the public sector by avoiding risks, whereas one mistake can kill a career. It also strengthens the tendency towards bureaucracy; rules provide a shield against criticism.” (OECD, 2009)

In any case, this **political will for rigorous reflection on what constitutes general interest and which policies will best serve it**, and to do everything possible to advance on this issue, is the fundamental driver for the start-up and development of evaluation. Without this political will, born of a **personal conviction** regarding the need to improve government action for the common good and the **intrinsic value of evaluation in this work**, the mere introduction of evaluation regulations, bodies and techniques in public administrations will hardly be enough to enable evaluation to put

down strong roots to prevent the system from crumbling and being irrecoverably weakened by any political change or new reform of the Administration.

In order to develop a successful evaluation system, that fundamental political will and personal conviction must be shared by many, not just among the political elite. These same convictions must be held by the public sector professionals responsible for managing the resources and programmes that provide services directly to the public, as well as by academic experts and civil society organisations, because the participation of people from all sectors will be needed in order for evaluation to develop.

Generating a political, social and administrative culture that serves as fertile ground for evaluation will therefore require time and resources to be invested in **raising awareness**, in **explaining** and **educating** all of these stakeholders on what public policy evaluation is and why it should be adopted in this continuous definition and pursuit of the common good. We are not referring to technical training on evaluation methodology, but to communication actions, seminars, working sessions, introductory workshops, conferences, meetings, etc. that provide a simple explanation of what evaluation is, the value it can bring and what it involves; where the results of evaluations can be presented in an appealing and understandable way and be debated among experts, managers, beneficiaries and members of the public.

Such actions for raising awareness and providing introductory training will also help in the adoption of common epistemological base. That is, a **shared understanding of what public policy evaluation is**, how it differs from other similar techniques described in earlier sections, what value it can bring and what is to be expected of it.

With this political will and shared cognitive universe as its foundation, the other elements seen in previous sections can be put together to build an evaluation system with a certain degree of sturdiness. But certain other contextual factors must also exist to provide suitable support to the process, as we saw in previous sections.

One such factor, which is fundamental, is **academia**, and generally all sources of expert knowledge on social issues, public policies, research methods and public policy evaluation. In order for public policy evaluation to develop, these sources of expert knowledge have to exist and reach the critical mass necessary to guarantee that the need for external evaluations can be suitably met.

But the mere **existence, critical mass and quality** of expert knowledge is not enough. This type of knowledge often suffers from excessive compartmentalisation in academic silos. Instead, evaluation, as we have seen, requires hybridisation, **multidisciplinary perspectives and teams**, and **mixed methods** research. Furthermore, the link between this expert knowledge and the administration is not apparent. The transfer of

knowledge from academia to the administration is neither sufficiently structured nor incentivised.

Administrations often forget to include in their calls for research grant applications specific initiatives aimed at answering questions with regard to the suitability of public interventions and gaining a more in-depth knowledge of certain social issues that occupy the top positions on political agendas. Moreover, academic careers based on high-quality scientific production, do not seem to offer sufficient incentives for knowledge transfer to the public sector. All too often there is a vast disparity between the time frames, questions, content and formats that meet the standards for scientific production and those suitable for the decision-making processes associated with public policy. Work will also need to be done in this area to establish **mechanisms that favour the transfer of knowledge and generally foster a more fluid relationship between the worlds of academia and public administration**, between expert knowledge and public policy decision-making processes.

In addition, work must focus on **promoting evaluation research beyond ad hoc projects, seeking to advance knowledge on social issues and to create analysis models and instruments for their evaluation**. There is therefore a need to stimulate, organise and financially support such research actions and activities for transferring know-how to the public sector.

As regards **formulas for institutionalisation**, we have seen that evaluation practices in government settings can be adopted in many forms. In fact, it can occur without necessarily requiring cross-cutting, specific organisations and regulations on evaluation, in environments where cultural factors encourage a regular evaluation practice without the need for such an institutional framework (e.g. in Germany). There can be a combination of favourable cultural and political factors, which in turn are linked to a certain degree of ad hoc regulation of evaluation and to support and continuous publication of evaluation practice through the use of incentives and capacity building actions (e.g. in the US Federal Government, which falls outside the scope of this study). Moreover, there are also favourable cultural contexts where evaluation has taken second place for a long time behind monitoring practices focussed on control (e.g. the United Kingdom). Furthermore, we have also seen cultural and political environments that remain unfavourable, where initial steps aimed at establishing formal institutions and regulations have not resulted in rapid advancement, at least for the moment (e.g. Belgium, Italy and Spain).

Nonetheless, **a certain degree of formal institutionalisation seems to be important** in order for evaluation to play any kind of role in government environments, particularly:

- As a way of providing it with visibility and transparency.

- As a way of protecting it, at least formally, against pressures from the political and bureaucratic system that could distort it (e.g. by proclaiming its independence or establishing an obligation to publicly disseminate evaluation results).
- To facilitate learning about evaluation, implementation methods and the connection with decision-making and accountability processes by establishing an explicit design and specific operating guidelines.

And a certain degree of formal institutionalisation may help also to **foster a predictable and regular evaluation practice** that adheres to **suitable quality standards** that promote its effective **use**.

We will now take a more detailed look at the different elements that form part of the institutionalisation process, including regulations, organisation and evaluation practice.

Firstly, in terms of **regulation** or the existence of specific laws or norms on evaluation, there does not seem to be a need for comprehensive legislation. This is usually covered, on one hand, by provisions included in cross-cutting and legally binding regulations governing public finances or national planning, and on the other, it is subject to internal guidelines published in orders, circulars or other documents setting out the evaluation policy. It is not as important for these cross-cutting regulations to govern the details of evaluation practice (which, as we have seen, is often carried out in a decentralised manner by the different government departments and agencies) but to impact some of its key elements: its objectives, criteria, quality standards, the resources assigned, how its results should be used and aspects associated with public dissemination.

Among the many regulatory instruments for institutionalisation of evaluation, in some countries we also saw the use of legislative clauses imposing evaluation obligations. Such provisions rarely go into detail and exist at a general level, opening the door to a range of implementation approaches.

It is clear that there is no single way to achieve institutionalisation through regulation, although experience does tell us that it is best to avoid an excessively legal approach that imposes poorly adapted evaluation obligations on unreceptive organisations or ones that do not have the technical capabilities to fulfil them, since this could lead to a risk of **increased bureaucratisation** of evaluation. Throughout this report, we have seen the diversity of evaluation institutionalisation processes, their highly idiosyncratic dynamics and their need for gradual and progressive implementation. In this context, there is nothing more counter-productive than confusing political support for the process with the mere imposition of new regulations on administrations that lack the capacity, the motivation or the necessary resources to undertake them.

In terms of the **organisational aspects**, in previous sections of this report we have seen various possible organisational configurations, which seem to be largely determined by the characteristics of the political systems to which they belong:

- A monopolistic configuration, where a single organisation has the dominant role in the promotion and practice of evaluation, with a predominantly regulatory and bottom-up approach. This configuration is seen in countries with a high level of centralisation and in the early stages of the development of evaluation. We have seen this in the case of the United Kingdom, not so much in relation to the evaluation system but rather in relation to the *Performance Measurement* system, in the early stages of its implementation. We also saw it in Chile, which clearly followed the Anglo-Saxon model in the beginning, although the system is currently much more flexible.
- A situation of 'centralised pluralism', in which one body occupies a predominant position, determining the behaviour of all the others but combining regulations and incentives to foster a more or less voluntary and gradual adoption of the system. We saw this configuration in the Netherlands, the United Kingdom, in Colombia, in France, and to a lesser extent in Mexico, where there are several forces driving evaluation which, in principle at least, act in a coordinated manner.
- A pluralistic scenario seen in highly decentralised countries such as Germany and Sweden.
- A fragmented concurrence scenario, seen in countries with poorly developed evaluation systems in their early stages, which aspire to a pluralistic or centralised pluralistic model but have not yet fully achieved it. The evaluation systems in Italy, Spain and Belgium seem to fit into this category.

In general, it is apparent that the institutionalisation of public policy evaluation takes many forms, which translates into a diverse range of agencies, department or groups of greater or lesser levels of complexity (public and private research institutes; central, regional or local administrations; autonomous agencies; the legislative branch; the executive branch; courts of audit; evaluation associations; private consultancy firms; philanthropic institutions; NGOs; etc.) at the central, regional or local level in each country, as well as at supranational level.

Within the administration it seems that the progress of evaluation is favoured by a **combination of bodies in charge of the centralised promotion, coordination and supervision** of evaluation, which also act as **resource and methodological support centres, with a decentralised evaluation function** adapted in a flexible manner to suit each organisational context.

Furthermore, the technical personnel within the administration who will manage or use evaluations must be equipped with the necessary **analytical and knowledge management skills**. Evaluation capacity cannot be easily improvised. Quickly

transforming civil servants working in the administration and with no background in research into expert evaluators is simply not feasible (although it is surprising to see how easily the opposite has been assumed), especially if we take into account the solid training they will need to receive on social research methods and techniques. It is not so much a question of training civil servants from the administration to become high-level expert evaluators (they will largely be found in the field of academia), but one of training evaluation promoters and managers, capable of converting a need for information and knowledge into questions that can be answered through evaluation, of acting as intelligent customers of externally commissioned services, of monitoring the use of evaluation reports and, where necessary, and more importantly, of preparing evaluation plans, ensuring the optimal internal evaluability conditions and creating a favourable internal culture. This is a specific function that should not be undertaken as an additional responsibility alongside other duties, without training, without resources and without a legitimate basis for doing so. On the contrary, great care should be taken with this strategic function charged with producing and managing quality knowledge to improve public policies. The **organisational functions associated with the evaluation system** must be **clearly defined** and **suitably positioned within the organisational structures and in the job catalogues** of public administrations. Suitable recruitment and professional development systems must be defined.

In terms of **evaluation practice**, it is important to accept that **progress on evaluation is gradual, irregular and asymmetric**. It therefore seems appropriate to promote **focal points for quality evaluation** wherever there is a combination of favourable factors and to later disseminate these ad hoc experiences among public sector organisations through **internal and inter-administrative networks**. These networks should foster a positive 'contagion' effect and a higher level of involvement and commitment to evaluation among public sector professionals. Throughout this report we have often seen evidence of this type of evaluation 'focal points' (such as PROGRESA evaluations in Mexico, or those associated with active employment policies in Germany, or with youth employment in France, etc.), alongside the creation of networks in practically all of the countries studied.

Promoting evaluation practice in public administrations also involves being seriously committed to improving the current evaluability conditions of policies, and, in particular, to the **availability and quality of the data contained in administrative records**, to its **use for statistical purposes** and to ensuring **easy access** to this information by evaluation professionals. The current weaknesses in this area often pose insurmountable problems for the execution of retrospective impact evaluations. In order to make progress on the necessary improvements, it is essential to secure the collaboration and coordination of strategies and actions among the agencies responsible for promoting evaluation, the departments in charge of official statistics, government agencies responsible for ICT and information systems, and institutions responsible for data protection. The current drive for open government and

e-administration policies offers a valuable opportunity to develop good administrative databases that can favour both evaluation and official statistics. However, the synergies between these initiatives are often ignored.

Planning evaluations in advance and adopting a prospective approach wherever possible, together with an adequate coordination of actions among the aforementioned bodies, would undoubtedly lead to a considerable improvement in evaluation conditions over the medium term. To this end, it would be useful to stimulate the creation of **annual or multi-annual evaluation plans or agendas**, a practice we have already seen in countries like Chile, Colombia, Mexico and the Netherlands, and also present in other countries outside the scope of this study (United States, Canada, Switzerland). These evaluation plans can be created by the agencies responsible for evaluation in centralised government settings, or by each individual government agency or department in decentralised settings.

The preparation of the evaluation plans or agendas requires **prioritisation**, for evaluation purposes, of a certain number of policies or programmes based on a set of criteria, which usually include aspects of political relevance (link to the government's strategic objectives, the sector of the population affected, etc.), budgetary aspects (budget allocated, queries with regard to cost-benefit or cost-effectiveness, etc.) and the viability of the evaluations (clarity of objectives and the design of the policy, availability of information and data, commitment of policy-makers and technicians of the programme to evaluation, etc.). Alongside these primary criteria, the evaluation agenda preparation process must also take into account other criteria associated with the characteristics of the programme to help determine the evaluation approach to be adopted (whether its a new programme or a mature initiative; results of previous evaluations, etc.).

These evaluation plans or agendas should therefore set out **which programmes are to be evaluated and what type of evaluation should be used in each case** (design evaluation, implementation evaluation, economic evaluation, retrospective impact assessment or prospective impact assessment). They should also include the **time line** for the evaluations, **assign the responsibilities** (which unit should drive and manage the evaluation), the **resources** (budget), and state the method of execution (internal or external, and the type of tender process to be used for external commissions, based on the budget).

Furthermore, evaluation plans should also include the **rules for public dissemination** of evaluations and the **mechanisms for monitoring the application of results and recommendations** (ideally echoing the provisions of a cross-cutting regulation in this regard).

The existence of well-designed policies, plans and evaluation agendas equipped with adequate economic and technical resources, together with the suitable evaluability conditions is not in itself sufficient to guarantee regular evaluation practice of quality. A key factor is still missing, which is the **demand** for evaluation from political authorities and/or the agencies who manage the programmes. Let us not forget that, in principle, the managing agencies and their leaders have no real incentive to want to subject their programmes to evaluation and be held accountable for their functioning. In addition, we have seen that highly centralised systems where the decision as to what, how and when to evaluate is imposed from above, lead to problems of a lack of buy-in which later result in poor levels of **use of evaluation results** for improving programmes. A technically high quality evaluation with results that are not used to improve policy is a useless evaluation. In order to avoid this problem, evaluation systems must include **incentives that stimulate demand and buy-in** for evaluation from programme managing agencies. Furthermore, wherever possible, these incentives must be positive and help create an internal culture and technical conditions favourable for evaluation, whilst also helping to provide access to the necessary financial resources. An example of this could be competitive grant funds that provide co-funding for evaluation projects, internal training activities and actions for improving evaluability conditions. This is without prejudice to the inclusion of any other type of incentives, positive or negative, particularly those associated with the allocation of budget resources.

It is also important to **involve all relevant stakeholders in the evaluation process**, including policymakers, programme managers, other public or private bodies involved in the design or implementation of the process, etc. This involvement serves various objectives: firstly, it helps to improve the level of buy-in for evaluation and therefore facilitates the subsequent use of the results in decision-making processes. Moreover, involvement of these stakeholders in the process also serves as an excellent opportunity to raise awareness and educate them on what evaluation is, what it can be used for, how it is carried out, etc.

In terms of evaluation practice, it is also important to adopt mechanisms that allow the **neutrality and quality of evaluations** to be guaranteed. This quality control can be organised in different ways, as we have seen throughout the report. For example, through groups or committees of external experts, created on an ad hoc basis for each evaluation, which certify the quality of the design and the execution of the evaluation, as well as the suitability of the results and recommendations. Alternatively, through committees with the same functions and characteristics, but created in a more stable format, to supervise all evaluations conducted, although the composition of the committees can be adapted to suit each case. Alternatively, quality control can also be exercised through the independent control and oversight agencies of the executive branch that regularly conduct meta-evaluations of the evaluations conducted or commissioned in particular areas of intervention. Also, the same function and activities can be carried out by multi-lateral bodies, international initiatives for the promotion of

evaluation or independent research centres. Public dissemination of evaluations and access to the databases used (with necessary data protection guarantees) can enable quality control to be carried out from multiple instances, including social entities.

Given that in this report we have focussed on a systemic approach to evaluation, merely ensuring quality control of each evaluation is not enough. Instead, we must also aspire to have **mechanisms capable of overseeing the overall functioning of the system** and introducing the necessary adjustments or corrections within the different elements to make them function properly. We have seen that in countries with greater tradition of evaluation, the bodies promoting evaluation in the executive branch or supreme oversight bodies periodically monitor the progress of evaluation from a global perspective, analysing how it is conducted, the resources it uses and, in particular, its impact in terms of improving and increasing the transparency of government actions. It seems important to include within this "systemic oversight", the continuous observation of new aspects in the configuration and functioning of evaluation systems, and in the policies and strategies implemented in the international arena, with a view to promoting them. As we have seen, this is a new area beginning to emerge in the literature on evaluation and public sector management. Specific research is needed on this issue, and so too is the continuity of the initiatives started by multi-lateral institutions such as the World Bank or the IDB to describe and analyse the functioning of such systems.

Finally, as stated in previous sections of this report, understanding what evaluation is, what are its assets and limitations, and committing to evaluation involves undertaking a process of **gradual and profound transformation of the current public administration model**. That transformation must overcome the strictly administrativist approach (centred on the necessary control of legality and the monitoring of the administrative procedure) and the managerialist approach (centred on the provision of quality services and on management control). It should be aimed at achieving a more effective public sector, able to cope with, and to mitigate, the variety of problems that require the intervention of public authorities. This requires an administration that is capable of supporting and managing the continuous creation and use of quality knowledge on such issues and interventions. If that is the administration that we want, we must, at some point, undertake an in-depth review of the current policies for administrative reform to ensure that priority is given to views and measures aimed at fulfilling that objective.

Appendix I. Bibliography

- ALCÁNTARA SÁEZ, M. (2013). *Sistemas políticos de América Latina*. Vol. I. América del Sur. Tecnos (4ª ed.).
- ARGILÉS, J.M. (2014). "Veinticinco años de evaluación en la política española de cooperación internacional para el desarrollo", en *Revista de Evaluación de Programas y Políticas Públicas*, No. 3, pp.19-52.
- ARENAS, A.; BERNER, H. (2010). *Presupuesto por Resultados y la Consolidación del Sistema de Evaluación y Control de Gestión del Gobierno Central*. Gobierno de Chile. Ministerio de Hacienda. Dirección de Presupuestos. Febrero 2010.
- BARBIER, J-C.; HAWKINS, P., (Eds.). (2012). *Evaluation Cultures. Sense Making in Complex Times*. Comparative Policy Evaluation, Vol. 19. Transaction Publishers. New Brunswick.
- HORBER-PAPAZIAN, K.; JACOT-DESCOMBES, C.; (2012), «Is Evaluation Culture Shaped by the Swiss Political System and Multiculturalism?», en BARBIER, J-C.; HAWKINS, P., (Eds.) (2012).
- SPEER, S. (2012), "Sectoral Evaluation Cultures: a Comparison of the Education and Labor Market Sectors in Germany", en BARBIER, J-C.; HAWKINS, P., (Eds.) (2012).
- STAME, N. (2012). "The Culture of Evaluation in Italy", en BARBIER, J-C.; HAWKINS, P., (Eds.) (2012)
- BASSANINI, F. (2009). *Twenty years of administrative reform in Italy*. <http://bit.ly/1aqphwY>
- BECK JØRGENSEN, T.; BOZEMAN, B. (2007). "Public Values. An Inventory". *Administration & Society*. Volume 39, No. 3, May 2007 354-381.
- BEMELMANS-VIDEC M-L., ERIKSEN B., GOLENBERG N. (1994). "Facilitating organizational learning: human resource management and program evaluation", en LEEUW F., RIST R., SONNICHSEN R., (Eds.), *Can Government Learn? Comparative Perspective on Evaluation and Organizational Learning*, New Brunswick, Transaction Publishers, 1994
- BOHNI NIELSEN, S.; WINTHER, D. M. (2014). "A Nordic evaluation tradition? A look at the peer-reviewed evaluation literature". *Evaluation*. Vol. 20(3) 311–331.
- BOYLE, R.; LEMAIRE, D.; RIST, R. (1999). "Introduction: building evaluation capacity", *Building Effective Evaluation Capacity: Lessons from Practice*, Transaction Publishers. New Brunswick.
- BRYSON, J; CROSBY, B.C.; BLOOMBERG, L. (2014). "Public Value Governance: Moving beyond Traditional Public Administration and the New Public Management". *Public Administration Review*, Vol. 74, Iss. 4, pp. 445–456.
- BUSSMANN, W. (2008). "The Emergence of Evaluation in Switzerland". *Evaluation*. Vol 14(4): 499 –506.
- CASADO, D. (2009). *Guía práctica 2. Evaluación de necesidades sociales*. Colección Ivàlua de guías prácticas sobre evaluación de políticas públicas. Instituto Catalán de Evaluación de Políticas Públicas (Ivàlua).
- CHELIMSKY, E. (2008). "A Clash of Cultures: Improving the 'fit' Between Evaluative Independence and the Political Requirements of a Democratic Society". *American Journal of Evaluation*, núm. 29, p. 400-415.
- CHELIMSKY, E. (2009). "Integrating Evaluation Units into the Political Environment of Government: The Role of Evaluation Policy, *New Directions for Evaluation*, núm. 123, p. 51-66.

- CHELIMSKY, E. (2014). "Public-Interest Values and Program Sustainability: Some Implications for Evaluation Practice". *American Journal of Evaluation*. Vol. 35(4) 527-542.
- COMMISSION OF THE EUROPEAN COMMUNITIES (2007). *Responding to Strategic Needs. Reinforcing the Use of Evaluation*. SEC (2007) 213. Brussels, 21 February 2007. <http://bit.ly/1A5P9ZG>
- CONEVAL (2013). *Diagnóstico para conocer el avance en monitoreo y evaluación en las entidades federativas 2013*. <http://bit.ly/1AYcyNz>
- CONEVAL (2014). *Informe de Evaluación de la Política de Desarrollo Social 2014*. <http://bit.ly/1u0970z>
- COOKSY, L. J.; MARK, M. M.; TROCHIM, M.K. (2009). "Evaluation policy and evaluation practice: where do we go from here", *New Directions for Evaluation*, núm. 123, pp. 103 – 110.
- CRÉMIEUX, L.; SANGRA, E.; (2014). *Evaluation at the Swiss Federal Audit Office*.
- CUNILL-GRAU, N., & OSPINA, S. M. (2012). "Performance measurement and evaluation systems: Institutionalizing accountability for governmental results in Latin America". In S. Kushner & E. Rotondo (Eds.), *Evaluation voices from Latin America. New Directions for Evaluation*, 134, 77–91.
- DELEAU, M., (1986). *Évaluer les politiques publiques: méthodes, déontologies, organisation: rapport*. Paris: Documentation française.
- DEPARTAMENTO NACIONAL DE PLANEACIÓN. (2010). *15 años del Sistema Nacional de Evaluación de Gestión y Resultados – SINERGIA*. DNP y Banco Mundial. Bogotá, D.C., Colombia.
- ESV (2013). *Financial information of good quality, in time and for multiple uses – the Swedish experience*. <http://bit.ly/1E0qdEr>
- EUROPEAN COMMISSION (2014). *Programming Period 2014-2020. Monitoring and Evaluation of European Cohesion Policy. European Social Fund*. September 2014. <http://bit.ly/1zWqEOZ>
- FEINSTEIN, O.; ZAPICO-GOÑI, E. (2010). "Evaluation of Government Performance and Public Policies in Spain". ECD Working Paper Series, núm. 22, May 2010. IEG. The World Bank Group.
- FORS, K.; REBIEN, C. (2014). "Is there a Nordic/Scandinavian Evaluation Tradition?". *Evaluation*, Vol. 20(4) 467–470.
- FURUBO, J-E. (2013). "When Incremental Change isn't Enough", en Rist, R., Boily, M-H, Martin, F.R (Eds.) (2013); *Development Evaluation in Times of Turbulence. Dealing with Crises that Endanger our Future*. The World Bank
- FURUBO, J.; RIST, R.; SANDAHL, R. (Eds) (2002). *International Atlas of Evaluation*. Transaction Publishers. New Brunswick.
- BEMELMANS-VIDEC, M.L. (2002). "Evaluation in the Netherlands 1990 – 2000: Consolidation and Expansion", en Furubo, Rist y Sandahl (Eds) (2002), págs. 91-114.
- DERLIEN, H-U., (2002); "Policy Evaluation in Germany: Institutional Continuation and Sectoral Activation", en Furubo, Rist y Sandahl, (Eds) (2002), págs. 77-91.
- FONTAINE, c.; MONNIER, E.; (2002) "Evaluation in France", en Furubo, Rist y Sandahl (Eds.) (2002) págs. 63-75.
- FURUBO, J.; SANDAHL, R.; (2002); "A Diffusion Perspective on Global Developments in Evaluation", en Furubo, Rist y Sandahl, (Eds.) (2002) págs. 1-23.
- FURUBO, J.; SANDAHL, R.; (2002); "Coordinated Pluralism – The Swedish Case", en Furubo, Rist y Sandahl, (Eds.) (2002), págs. 115-128.
- GALLEGO, R. (2013). "Decision-making and policy shift under quasi-federal constraints: The case of health management in Catalonia, 2003-2007". Paper presentado en la *63rd Political Studies Association (PSA) Annual International Conference*. 25 - 27 Marzo 2013, Cardiff.
- GALLEGO, R.; SUBIRATS, J. (2012). "Spanish and Regional Welfare Systems: Policy Innovation and Multi-level Governance". *Regional and Federal Studies* Vol. 22, No. 3, 269–288, July 2012.

- GURGAND, M., VALEDENAIRE, M. (2012). «Le Fonds d'Expérimentation pour la Jeunesse», en *Éducation & Formations* no. 81. <http://goo.gl/mwKuM>
- GUZMÁN, M.; IRARRÁZAVAL, I.; DE LOS RÍOS, B. (2014). "Monitoring and Evaluation System: The Case of Chile 1990–2014". ECD Working Paper Series, No. 29, July 2014. IEG. The World Bank Group.
- HERNÁNDEZ LICONA, G. (2009). *Evaluación de la Política Social: Un cambio institucional en México*.
- HERNÁNDEZ LICONA, G.; MARTÍNEZ MENDOZA, E.; GARZA NAVARRETE, T. (2013). *Construyendo un sistema de evaluación del desempeño para el desarrollo social*.
- HOLM, G. (2006). "Evaluation in the Nordic Countries". *Journal of Multi-Disciplinary Evaluation* (JMDE:4)
- INNERARITY, D. (2015). "La decepción democrática", diario *El País*, 2 de febrero.
- JACOB, S. (2005). *Institutionaliser l'Évaluation des Politiques Publiques. Étude comparée des dispositifs en Belgique, en France, en Suisse et aux Pays-Bas*. P.I.E. Peter Lang.
- JACOBI, L.; KLUVE, J. (2006). *Before and After the Hartz Reforms: The Performance of Active Labour Market Policy in Germany*. Discussion Paper Series. IZA. April 2006.
- JACOB, S.; SPEER, S.; FURUBO, J-E. (2015). "The institutionalization of evaluation matters: Updating the International Atlas of Evaluation 10 years later". *Evaluation* 2015, Vol. 21(1) 6–31.
- KICKERT, W. (2011). "Distinctiveness of administrative reform in Greece, Italy, Portugal and Spain. Common characteristics of context, administrations and reforms". *Public Administration*, Vol. 89, No. 3, 2011 (801–818).
- KUHLMANN, S.; BOGUMIL, J.; GROHS, S. (2008). "Evaluating Administrative Modernization in German Local Governments: Success or Failure of the New Steering Government?". *Public Administration Review*; Sep/Oct 2008; 68, 5; pg. 851.
- KUHLMANN, S. (2010). "Performance Measurement in European local governments: a comparative analysis of reform experiences in Great Britain, France, Sweden and Germany". *International Review of Administrative Sciences*, Vol. 76(2):331–345.
- LÁZARO, B. (2012a). "Estrategias para el avance de la evaluación: la experiencia catalana a través de Ivalua". *Presupuesto y Gasto Público*. 68/2012: 25–39. Secretaría de Estado de Presupuestos y Gastos. Instituto de Estudios Fiscales.
- LÁZARO, B. (2012b). "Las nuevas formas de inversión social como motor de innovación de las políticas públicas", en *Avaluació per al Bon Govern*, Noviembre de 2012. <http://bit.ly/160vYTJ>
- LEEuw, F.L. (2009). "Evaluation Policy in the Netherlands". *New Directions in Evaluation*. Special Issue: Evaluation Policy and Evaluation Practice. Issue 123, pages 87–102.
- LEEuw, F.L.; FURUBO, J. (2008). "Evaluation Systems. What are They and Why Study Them?" en *Evaluation*. Vol. 14(2), pp. 157–170.
- MACKAY, K. (2007). *How to Build M&E Systems to Support Better Government*. The World Bank.
- MARK, M.M.; COOKSY, L.J.; TROCHIM, M.K. (2009). "Evaluation Policy: An Introduction and Overview", *New Directions for Evaluation*, núm.. 123, pp. 3–12.
- MAY, E.; SHAND, D.; MACKAY, K.; ROJAS, F.; SAAVEDRA, F., (Eds), (2006). *Hacia la institucionalización de los sistemas de monitoreo y evaluación en América Latina y el Caribe*. Banco Mundial. Banco Interamericano de Desarrollo. Washington D.C.
- NATIONAL AUDIT OFFICE (2013). *Evaluation in Government*. Cross-government Report. December 2013. United Kingdom.
- NEIROTTI, N. (2012). "Evaluation in Latin America: Paradigms and practices". In S. Kushner & E. Rotondo (Eds.), *Evaluation voices from Latin America*. *New Directions for Evaluation*, 134, 7–16.

Consortium led by



Coordinating Partners



Over 80 Participating Operational Partners and Collaborating Institutions in Europe and Latin America

EUROsocial is a regional cooperation programme between the European Union and Latin America for the promotion of social cohesion through support for national public policies and the strengthening of the institutions that put them into practice. EUROsocial aims to promote a European-Latin American dialogue about public policies surrounding social cohesion. Its aim is to contribute to reform and implementation processes in ten key areas of public policy in certain thematic areas selected for their potential impact on social cohesion. The instrument it provides is institutional cooperation, or peer-to-peer learning, meaning the exchange of experiences and technical advice between public institutions of Europe and Latin America.

