Evaluation Framework and Practices: A comparative analysis of five OECD countries

Axel Mathot, Senior Policy Analyst, OECD

Flavia Giannini, Junior Policy Analyst, OECD

This article provides a detailed analysis of policy evaluation frameworks in five OECD countries (Canada, Germany, the Netherlands, the United Kingdom and the United States), with a focus on the institutionalisation, the quality and the use of evaluation across government. It discusses how the evaluation function is regulated and organised in these countries. In addition, the article provides insights on the relevant conditions to set up and develop a sound evaluation system, through capacity building activities, quality assurance and control mechanisms, data availability and the use of evaluations.

JEL classification: H11, H50, H83

Keywords: evaluation, transparency, public finance

Acknowledgements

This publication was authored by Axel Mathot and Flavia Giannini in the Public Management and Budgeting Division (PMB). The authors would like to express thank their colleagues from PMB who contributed to this paper by providing a structured overview of the framework and practices in Canada (Stéphane Jacobzone), United Kingdom (Camila Vammalle) and United States (Andrew Blazey). Claire Salama and Andrew Blazey provided useful comments throughout the drafting. Peter Van Humbeeck contributed to the first draft of the synthesis paper. The authors are also grateful to the many people in – and outside the administrations of the five countries who shared their insights on the topic.

1.1. Introduction and key findings

The analysis of a policy evaluation framework requires a systemic view that takes into consideration the institutionalisation, quality and use of this practice. Policy evaluations are not a standalone process: they are – or should be – integrated within the whole policy cycle to pursue different objectives, such as management and policy improvements, greater accountability and evidence-informed decision making.

This article summarises the findings from five case studies of countries with policy evaluation systems that follow good practices in the design, conduct and use of policy evaluations: Canada, Germany, the Netherlands, the United Kingdom and the United States. Although the case studies covered a very broad range of evaluation activities (*ex ante*, performance monitoring, spending reviews and *ex post* evaluations), this synthesis note focuses on *ex post* evaluations.[1]

This benchmark exercise builds on previous OECD work, mainly the publication “Improving Governance with Policy Evaluation, Lessons from country experiences” (OECD, 2020[1]) which offers a cross-cutting contribution to the global policy debate on evaluation and evidence-informed policy making, relying on a survey of 42 OECD and non-OECD countries.

The comparison of the policy evaluation framework and practices in Canada, Germany, the Netherlands, United Kingdom and the United States, show that there is no one-size-fits-all approach to set up an evaluation framework. Based on the experience of these five countries, it appears that building an evaluation culture requires a whole-of-government approach that needs to take into account the local political and institutional context. The culture and practices of a country can in fact determine key governance choices around a policy evaluation system, for instance whether to establish a framework in the legislation or keep it rather informal, or whether to allocate evaluation responsibilities to the centre of government or grant more discretion to the line ministries. Each option has its advantages and may have its drawbacks, but ultimately depends on the prevailing traditions in a given public administration, which explain the diversity of approaches to policy evaluation among OECD countries, and, in particular, among the countries selected in this study.

Despite the heterogeneity of how the policy evaluation function is organised across countries, several conditions can be identified for the development of a robust evaluation system. First, building capacity to carry out evaluations at the ministry level is essential to conduct good quality analysis and also to avoid outsourcing all evaluations to external actors, which however can still be useful to guarantee independence and perform a quality check. In-house capability and knowledge should be leveraged to generate relevant outcomes from evaluation, which will be useful for internal learning purposes. Secondly, to be effective, a sound framework should be supported by the regular use of policy evaluation results, including their integration into the budget or having clear guidance on the follow-up of evaluations. Promoting public availability and accessibility of evaluations not only ensures transparency but also stimulates the use of evaluation findings, both by the executive and by the Parliament. The latter can in fact play a major role to make the policy evaluation system responsive and impactful, in that it can initiate, demand, commission, use or even produce evaluations.

Finally, creating a proper eco-system in which the policy evaluation apparatus can flourish is key to its successful implementation and continuous improvement. This can translate in the production of evaluation guidelines and standards, the involvement of relevant stakeholders, and the availability of high-quality data, which is of utmost importance for conducting evaluation activities but still represents a challenge in several countries.

This article is divided into four sections:

1. The first section (1.2) provides the definition of policy evaluation developed both by the OECD and the countries under study.
2. The second section (1.3) elaborates on crucial elements of policy evaluation systems (focusing on governance) and answers key questions such as: What does a sound policy framework look like? Is a legal basis necessary and what should it cover? Which institutions develop and adjust the framework and provide oversight, co-ordination and guidance? What policies to evaluate and when? What is the role of parliament and supreme audit institutions?
3. The third section (1.4) discusses specific questions around the policy evaluation process: How to conduct evaluations? Who conducts the evaluations? How to create transparency and involve stakeholders?
4. The fourth section (1.5) dives into the conditions to develop a robust policy evaluation culture: How to build capacity? How to ensure quality? How to increase data and information availability? How to promote use of evaluations?

1.2. On the definition of policy evaluation

1.2.1. Definition of policy evaluation as part of the policy framework

The OECD uses the following working definition of policy evaluation: "the structured and evidence-based assessment of the design, implementation or impacts of a planned, ongoing or completed public intervention" (OECD, forthcoming[2]). This definition takes a broad view on the type of evaluations and points to the usefulness of this tool in supporting evidence-informed policies.

The main features of the OECD definition can be found in national definitions, but there are some differences in how countries phrase it:

- In Canada, policy evaluation is defined as the systematic and neutral collection and analysis of evidence to judge merit, worth or value, which typically focus on programmes, policies and priorities and examine questions related to relevance, effectiveness and efficiency. The purpose is to inform decision making, improvements, innovation and accountability (Treasury Board Canada, 2016[3]).
- In Germany, policy evaluation is defined as a procedure that compares the originally formulated expectations regarding objectives, benefit and costs with the actual effects, unintended consequences and costs. It is based on a systematic methodology, a process that can easily be understood by third parties, and precise empirical data sets (NKR, 2018[4]).
- In the Netherlands, policy evaluations are examinations of the effectiveness and efficiency of policies (art. 4.1 of the Accountability Act).
- In the United Kingdom, policy evaluation is defined as a systematic assessment of an intervention's design, implementation and outcomes, where an intervention could be either a policy, a programme or any other government activity meant to obtain a change (UK HM Treasury, 2018[5]).
- In the United States, evaluation is an individual, systematic assessment of one or more programmes, policies and organisations, using systematic data collection and analysis and intended to assess their effectiveness and efficacy. Evaluations may address questions related to the implementation of a programme, the effectiveness of programme strategies, or factors that relate to variability in effectiveness of the programme or strategies (OMB, 2020[6]).

All the definitions in the benchmark countries are broad, but the analytical aspect is included explicitly in most of them. As for the objective of policy evaluation, this is defined in a non-specific way in the case of the Netherlands (only referring to efficiency and effectiveness), while the other benchmark countries are more specific, always including the aspect of efficiency and effectiveness. Both Canada and Germany refer explicitly to the policy cycle and the fact that evaluations are linked to the decision-making process and thus policy recommendations.

1.2.2. Complementary instruments and practices

The focus of this paper is on *ex post* policy evaluation. There are similarities between *ex post* policy evaluation and other instruments that can support a wider evaluation framework, for example:

- *Spending reviews* are tools for developing, assessing, recommending and adopting policy options by analysing the government's existing expenditure, and linking them to the budget process. The purposes of spending reviews can be to enable the government to manage the aggregate level of expenditures, to align expenditures according to the priorities of the government or also to improve the effectiveness within programmes and policies (OECD, forthcoming[7]). Evaluations are an important part of spending reviews and the budget process as they contribute to the assessment of expenditure.
- *Performance* monitoring aims to facilitate planning and decision making by providing evidence to measure performance and help raise specific questions to identify implementation delays and bottlenecks. It can also strengthen accountability and public information, as information regarding the use of resources, efficiency of internal processes and outputs of policy initiatives is measured and publicised. Unlike evaluation, monitoring is driven by routines and ongoing process of data collections (OECD, 2021[8]). Thus, performance monitoring provides information about the current status of a programme or policy and the results of monitoring can inform the topics that are selected for evaluations.
- *Audits* determine whether the information collected or actual conditions correspond to established criteria, including compliance with financial or legal rules. Auditing helps to ensure that public-sector entities and public servants perform their functions effectively, efficiently, ethically and in accordance with the applicable laws and regulations (International Organisation of Supreme Audit Institutions, 2019[9]).

1.3. Governance matters

1.3.1. Legal bases of public policy evaluation frameworks

A policy framework is generally a document or a set of documents that provide strategic direction, guiding principles and courses of action to the government for a specific sector or thematic area. Many countries have specific policy frameworks to allocate institutional responsibilities for evaluation of public policies. These frameworks are often translated in a legal text.

A legal basis to undertake policy evaluations can be helpful, but is not necessary nor sufficient to ensure the quality of the findings and the effective use of evaluation results. The existence of a sound legal framework can be important to promote policy evaluation and clarify institutional responsibilities from a legal perspective (OECD, 2020[1]). As for the level at which policy evaluation is legally embedded, some benchmark countries have specific stipulations in the constitution. This is the case in Germany, where Article 104b states the necessity to conduct evaluations of financial assistance grants on a regular basis. Canada, the Netherlands and the United States define their evaluation framework in primary and/or secondary legislation and complement this legislation with guidelines, which provide more detailed requirements on evaluation annual plans, timing, selection criteria, etc. In these countries, primary legislation is used mainly to state the principles and importance of evaluations.

On the other hand, the United Kingdom has no legal framework for policy evaluation. The absence of legislation has not per definition proven to be detrimental for the conduct of evaluation in this country and in other countries with no legal framework. Hence, having a legal framework guiding policy evaluation across government is not a condition for success. The context and culture embedded in the administration play a considerable role. The risk when there is no legal framework supporting public policy evaluation, is that initiatives may be discontinued and the implementation and use of evaluations depend on the importance given by the current leadership. A key challenge in such non-institutionalised frameworks is to ensure the continuity of the evaluation framework across different governments, to avoid that the efforts to build an evaluation culture are reverted.

The legal framework balances between clarity and flexibility. The experience in the 5 benchmark countries suggests that existing regulatory/legal frameworks only define the basic evaluation obligations of government departments and agencies, and are complemented by more detailed guidelines and planning requirements.

Besides the existence of a legal framework, an important question is whether there is a policy framework in place. Policy frameworks can include different legislative acts, such as in the Netherlands and in the United States (OECD, 2020[1]). This is not necessarily the case, however, and some countries adopt evaluation policies in the form of guidelines without specific legislation.

- In Canada the policy evaluation framework is part of the Policy on Results, issued by the Treasury Board Secretariat pursuant to the Financial Administration Act (art. 2.1. Policy on Results). Although there is a clear foundation for the evaluation framework, it is not part of the Act itself. Paragraph 42.1 (1) of the Financial Administration act requires to evaluate grants and contributions programmes every five years on a rolling basis (Government of Canada, 1985[10]).
- The United States constitutes an interesting example. The Foundations for Evidence-Based Policymaking Act of 2018 mandates evidence-building activities across agencies and open government data, with the Office of Management and Budget (OMB) mandated to develop guidance and advice on policy evaluation. Although the Evidence Act elevates programme evaluation as a key agency function and gives guidance on the requirements for each entity, it also allows for tailor-made solutions within the different departments.

1.3.2. Organising the policy evaluation function across government

There is no single recipe for institutionalising policy evaluation across government. Some countries have centralised systems with one (or several) government department/agency co-ordinating evaluation activities across government, while other countries have more decentralised approaches. Factors such as the political system, public administration cultures and the rationale for evaluation shape the development and characteristics of the evaluation function and consequently how this is organised across government (OECD, 2020[1]).

A centralised system is more likely to result in a common evaluation approach promoted across government, where different ministries adopt similar practices with regard to the planning, implementation and use of evaluations. A degree of centralised co-ordination is also a prerequisite for a strategic approach to evaluation, which entails the ability to look across a number of evaluations that cover different aspects of the same policy area to draw overall conclusions and identify common learning points.

A more decentralised approach to organising the evaluation function makes it easier for individual ministries to customise evaluation methods and practices to their own specific requirements. The drawback is that evaluation may be neglected or undertaken in a way that does not benefit from the sharing of experience and know-how. Such a fragmented approach may also make it more difficult to adopt a strategic perspective to evaluation or to aggregate evaluation results across different parts of government that share common policy objectives.

All the five selected countries have established a hybrid system, which combines central steering with some degree of autonomy to organise the evaluation function within the line ministries/departments, both regarding the institutional set-up and the way evaluations are conducted.

Central steering vs. autonomy of line ministries

The balance between central steering and autonomy for the ministries/agencies varies a lot between the five countries. An overview of how evaluation is organised in the countries under study in terms of centralised and decentralised approaches is provided below. Rather than using the alphabetical order, the countries are listed according to their degree of centralisation, from highest to lowest.

United States

The Evidence Act in the United States is an example of central steering since it streamlines the evaluation system in the different departments (US Congress, 2019[11]). Pursuant to the Act, agencies are required to:

- Submit an evidence-building plan (referred to as Learning Agenda), included in the departmental or agency strategic plan to identify and address policy questions relevant to programmes and policies.
- Prepare an annual Evaluation Plan that describes the key evaluation activities that the department or agency plans to conduct in the next fiscal year.
- Appoint an Evaluation Officer to oversee evaluations and related activities and implement the agency evaluation policy; this Officer represents a clear point of contact for OMB and plays a clear role regarding the performance of the evaluation system in a federal department or agency.

Canada

The Canadian system has also a relatively strong co-ordinated overall framework. The President of the Treasury Board of Canada has been given a mandate to instil a culture of evaluation, measurement and evidence-informed decisions in programme and policy design and delivery. The Policy and Directive on Results (both of 2016) outline requirements for the departments such as the identification of roles and responsibilities for federal departmental officials. Every department must have a Head of Evaluation and a Performance Measurement and Evaluation Committee (PMEC), and larger departments must prepare a formal evaluation plan or conduct an evaluation planning exercise to identify the evaluation needs. Accountability for the quality of policy evaluation rests with the departments and the Treasury Board is not mandated to directly challenge the evaluations done within the departments. Departments are given flexibility in the design of their evaluations, but are also subject to central guidance under the Standard on Evaluation (Directive on Results, Appendix C) that, among other things, requires that evaluations are planned taking into account: the risks and complexity associated with the policy, programme, priority, unit or theme being evaluated; consideration of using relevance, effectiveness and efficiency as primary evaluation issues, where relevant to the goals of the evaluation; and government-wide policy considerations where relevant, such as gender-based and diversity analysis and official languages.

The Netherlands

The Dutch system is less centralised. The organisation of the policy evaluation function within Ministries is left to the responsible Minister. The Regulation for Periodic Evaluation stipulates that every policy must be evaluated on a regular basis (7 years) to examine its effectiveness and efficiency. Each ministry has to establish a directorate for Financial and Economic Matters (FEZ) to co-ordinate and supervise the evaluation on effectiveness and efficiency of policies, among other tasks. Although the FEZ do not conduct evaluations themselves, they are a focal point for the evaluations within the line ministry. Within this framework it is up to each Minister to comply and organise the evaluation function as they see fit. The Regulation does not impose a specific organisation in charge of carrying out the evaluations, and thus can be considered slightly more decentralised than the Canadian and US examples. The recent evolutions indicate that there is a tendency for more central guidance as a result of the operation Insight in Quality, which introduces new instruments such as the Strategic Evaluation Agenda and the Public Value Scan.

United Kingdom

A similar approach is implemented in the United Kingdom. Each department is expected to follow the central guidelines but has the flexibility to adapt them to its specific needs as long as they comply with basic principles. Within the overall framework, each department is responsible for developing its own policy evaluation strategy and evaluating its own policies.

Germany

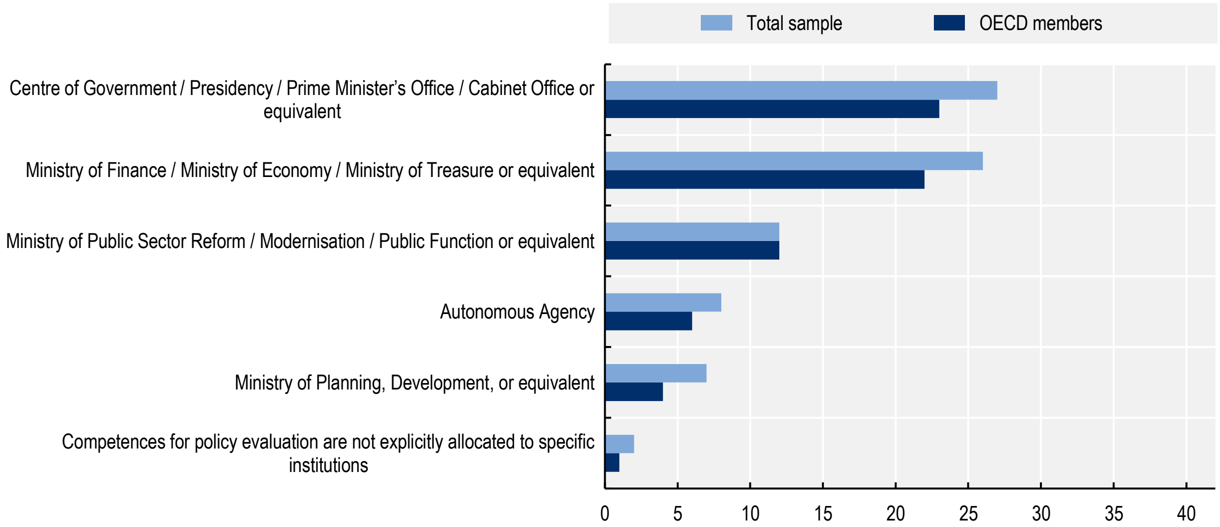
The most decentralised system can be found in Germany. Due to the so-called 'department principle', which states that each federal minister manages its area of competence independently and under its own responsibility, there is no uniform approach to evaluations and the internal organisation of the evaluation function varies widely among ministries. An example of the diversity is that some have their own specialised research institutions. This is the case for the Federal Office for Economic Affairs and Export Control, which has had an evaluation unit since 2008, and the three federal ministries for Economic Affairs and Energy, of Family, Senior Citizens, Women and Youth and of Education and Research.

The German case seems to illustrate that the absence of central oversight implies some challenges. The Court of Auditors, which verifies that evaluations conducted by federal ministries are appropriate and comply with section 7 of the Federal Budget Code, regularly signals gaps in terms of methodology, competences and resources for evaluation. The National Regulatory Control Council (NKR) highlighted that evaluations carried out in 2017-18 did not use clear evaluation criteria nor empirical analysis, and did not provide concrete recommendations (NKR, 2018[4]). Recent evolutions have shown that there was a need for a certain degree of central steering. In 2013, a systemic framework for *ex post* evaluation was established for important legislative initiatives. This framework introduced a threshold to decide whether an evaluation should be conducted, but did not include requirements that evaluation should be conducted. The 2013 concept for evaluation was updated in 2019, when a decision of Secretaries of State clarified additional aspects to be included in the draft bills that would facilitate the conduct and improve the quality of *ex post* evaluations. When presenting a draft bill, the objectives of the bill and the criteria for achieving the objectives have to be stated, as well as the steps and methods of evaluation that will be used. Furthermore, the Federal Government is developing a guidance document outlining the steps and methods of an evaluation (BPA, 2019[12]).

Allocation of responsibilities for developing the framework and providing oversight, co-ordination and guidance

Evidence from OECD countries shows that usually there are several institutions with competences for policy evaluation across government, which co-operate or have separate competences in the field.

Figure 1. Institutions with competences related to policy evaluation across government



Note: n=42 (35 OECD member countries). Answer option "other" is not displayed.
Source: OECD (2020) Improving Governance with Policy Evaluation.

As highlighted in the previous section and shown in Figure 1, most countries have chosen some form of central steering. Governments have established dedicated units and/or given the competency to a ministry to champion an evidence/evaluation agenda across government. They issue guidelines, stimulate capacity building, ensure quality of evaluations and manage an evaluation calendar. The central institution (or oftentimes a combination of institutions) has a key role in managing the evaluation eco-system, making sure that they can take place at the right time and in the right place and feed into decision making. In addition, these central units usually have the political leverage to ensure that the findings of evaluations are subsequently used. Therefore, allocating this central steering role to an institution close to political power can be interpreted as a sign of political commitment.

In each of the selected countries, with the exception of Germany, there is at least one central actor responsible for organising and co-ordinating the evaluation system. Table 1 lists the main institutions responsible for developing the policy evaluation framework and providing central steering and guidance. These entities do not conduct policy evaluations themselves, as this responsibility lies within the line ministries/departments.

Table 1. Main actors responsible for the evaluation framework in selected countries

Benchmark country	CoG	MoF	Other	Roles of the main responsible actors
United States	✓		✓	The Office of Management and Budget, within the Executive Office of the President is the main actor. Additional competences for evaluation across government reside in the Office of Evaluation Sciences within the General Services Administration.
The Netherlands	✓	✓		The Ministry of Finance has the main responsibilities for the evaluation framework but all ministries are responsible for conducting the evaluation of the policies in their field.
Canada	✓	✓		• Treasury Board Secretariat (TBS): the TBS-led process tracks permanent results and reports to support accountability and effectiveness in the medium term. This process aims to clarify the financial aspects and the resources that went into programmes and what was achieved. • Privy Council Office (PCO): The Results and Delivery process in the PCO focuses on shorter-term political mandates. The fact that there are two central entities involved creates the need for co-operate. First, under the guidance of the PCO, a new governance body called “Central Agencies and Justice” has been created. This body consists of the Ministry of Finance, the Ministry of Justice (in charge of reviewing the legal aspects of new legal proposals), the PCO, and the TBS. The objective of this new governance body is to share agendas and discuss priorities, establishing a starting point for high-level co-ordination. Secondly, the PCO and TBS are working to identify how the tools elaborated by the PCO could be leveraged by the TBS as conditions for funding to departments.
United Kingdom	✓	✓	✓	Her Majesty’s Treasury (HMT) plays an important role through the spending review process. The evaluation task force, a joint Cabinet Office-HMT unit, supports and challenges departments on their use of evaluation evidence, prioritisation and evaluation design. It also provides departments with access to external advice and evaluation support via the Trial Advice Panel and the What Works Network. The Cross-Government Evaluation Group (CGEG) aims to improve the supply of, stimulate demand for, and encourage the use of, good quality evaluation to inform the policy development process in order to achieve better outcomes for the public. They also produced the Magenta Book (guidance on evaluation).
Germany	✓	✓	✓	Actors with some responsibilities regarding *ex post* evaluations are present both within the Ministry of Finance and other central agencies. In the Centre of Government, the Federal government is the co-ordinator for better regulation; the National Regulatory Control Council (NKR) is an independent body at arm's length.

Note: CoG = Centre of Government; MoF = Ministry of Finance.
Source: Authors and (OECD, 2020[1]).

1.3.3. Setting the policy evaluation agenda: Periodicity and scope

Evaluations are time consuming and costly. In the OECD the majority of countries developed a selection process or criteria for determining when and what type of evaluation is needed (OECD, 2020[1]). Three practices seem to emerge:

- Establishing formal rules for periodic evaluations or for evaluation of certain policies (including supranational requirements such as EU requirements for evaluation of structural funds).
- Setting threshold and proportionality criteria.
- Installing a planning process to define and translate ambitions into strategic and operational evaluation plans or agendas.

The two main elements to determine are periodicity and scope (including the planning process), and the selected countries have taken different decisions on how to organise them.

Canada

The framework of the Policy on Results foresees a comprehensive coverage of all programme spending, which should be evaluated approximatively every five years, but the Policy on Results allows in reality for flexibility in conducting evaluations and their scope. The Directive on Results states that in the large departments a five-year evaluation plan has to be submitted prioritising the evaluations based on programme risks, governmental and departmental priorities and the information needs of major stakeholders. In addition, the Financial Administration Act stipulates the requirement to evaluate grants and contributions programmes every five years on a rolling basis (Government of Canada, 1985[10]). In order to organise the programme of evaluation, the evaluation directorates identify the mandatory requirements for evaluation and then add the needs identified by the managers. Given the costs of evaluation, it is also important to identify what can be done by the programme managers and front-line units in terms of “light analysis”, while formal evaluations are subject to a thorough process and inclusion in the 5-year rolling plan. The agenda is set internally in each department but every fall there are consultations with both the officers and team members of large departments and the TBS. The TBS suggests aspects to be looked at, and these consultations have an impact on the final choice.

Germany

Since 2013 there is an objective selection criterion stipulating that policies that have a considerable financial impact for citizens or the economy have to be evaluated every 3 to 5 years. *Ex post* evaluations are mandatory for ‘major’ regulatory projects, namely when *ex ante* estimates predict at least EUR 1 million of compliance costs for the economy and businesses or 100.000 hours of compliance time for citizens. Besides this mandatory evaluation, it is up to the ministries to decide on the evaluations they want to conduct, for example considering high total financial costs, particular political significance or great uncertainty on the proposal’s effects (State Secretaries Committee on Bureaucracy Reduction, 2013[13]).

The Netherlands

Mandatory evaluation is prescribed for all policies over a period of 7 years, or 5 years for subsidies. The operation Insight in Quality has revealed that mandatory evaluation of each policy is not necessarily the best solution. A mandatory evaluation within a clear timeline does not per definition provide the information that is needed by policy makers and the timing is not necessarily co-ordinated with the political agenda. Therefore, the Netherlands has introduced the Strategical Evaluation Agenda (SEA) with the aim to generate more relevant insights at the right moment. The SEA offers the opportunity for line ministries to organise the evaluation framework on a broad thematic basis (not necessarily limited to a specific policy) and to consider when it is strategically useful to plan an evaluation (Ministerie van Financiën, 2021[14]). The departmental SEA are sent to the Parliament annually and look 3-4 years forward. They give an overview of the key policy themes and prioritised evaluations (*ex post* as well as *ex ante* and *durante*). This allows the Parliament to have a better overview of the evaluations that will be conducted. There are benefits in reviewing policies as a group, rather than in a piecemeal fashion, where they are interactive or operate jointly to achieve related policy objectives. This implies some form of co-ordination and consultation.

United Kingdom

In the United Kingdom the resources and effort employed in monitoring and evaluation should be related to the scale of the proposals under consideration according to the guidelines of the Treasury (UK HM Treasury, 2018[5]). Thus, in practice, not all policies are evaluated. Government departments prioritise which policies to evaluate, including: policies that have formal requirements to be evaluated (i.e. evaluation clause introduced in the law), policies considered financially important or particularly emblematic of the government’s political line, and policies with potential learning benefits. In other cases, the evaluation of some policies and programmes is prompted by the National Audit Office or Parliamentary reports and recommendations.

United States

In the United States, departments determine the scope and periodicity of evaluations when preparing their evaluations plans. In some cases, Congress mandates that a particular evaluation occur for a certain programme within a certain timeframe. The Evidence Act introduced some innovations, such as the Learning Agenda to identify and address priority questions relevant to programmes and policies of a department or agency. However, such innovations do not require legislation on scope or timing, as these are decided by departments and agencies.

1.3.4. Learning and accountability

Policy evaluation serves not only accountability and better decision making at the central level, but should also have a positive impact on the quality of decision making in the long run, and allow for opportunities to improve the internal processes which can be complicated by focusing solely on accountability. The institutional framework for evaluation can include some elements that stimulate the learning aspect. For instance, the location of the central entity responsible for evaluations within the government can have an impact on this balance. In the Netherlands, the fact that all regulation and guidelines emanate from the Ministry of Finance implies that there is (at least organisationally) a link with the budget process. This institutional set-up may, at times, tilt the balance in favour of conducting evaluations for accountability rather than learning. This is something that the Ministry of Finance has identified and is seeking to remedy.

Two country examples stand out because of the importance attached to internal learning from evaluations and evidence-informed policy making in general:

- The operation Insight in Quality in the Netherlands intended to bring about a shift of attention from accountability towards a learning-oriented perspective. The introduction, inspired by the UK experiences, of a Public Value Scan, is a good example in this respect. Rather than trying to find an answer to the question of whether a policy has been efficient and effective, it focuses on whether everything is done to maximise the expected added value of the policy under the current circumstances, with input from various experts, relevant stakeholders and the policy officers within the ministry. The focus is not primarily on the results and outcomes achieved (product evaluation), but rather on the activities undertaken and efforts made (process evaluation). Furthermore, the emphasis is on the current policy rather than on an *ex post* justification of the policy pursued up to that point. This means that the Public Value Scan can be considered as a form of *durante* evaluation. This instrument is very new (only one pilot has been conducted for the moment) and its relation with the other existing tools, such as the policy reviews,[2] has still to be determined. The innovations introduced by the Operation Insight in Quality go beyond the Public Value Scan. The so-called “Improvement paragraph” in Policy Reviews requires ministries to include a paragraph where they explain what they will do to fill the information gaps and thus gain a better understanding of the efficiency and effectiveness of policies for future reviews. Similarly, the Strategic Evaluation Agenda aims to gather more relevant insights in due time, but also to facilitate learning, as the planning and preparation of evaluations increase the programme managers’ understanding of their programmes. Noteworthy in this respect is a 2018 initiative of the Ministry of Finance that started awarding the prize for ‘Best grant evaluation’. The prize is intended to draw attention to the quality of these evaluations, to highlight the positive points and provide tips to improve the quality. The jury includes representatives of the Court of Audit, the Netherlands Bureau for Economic Policy Analysis, the Bureau for Statistical Analysis and the Ministry of Finance.
- In the United States, the OMB oversees the implementation of the Evidence Act. The Act requires that evaluation is integrated into the existing strategic planning processes of a department and within its existing organisational structures. An important element is that the evaluation work should be embedded in the administrative procedures. The Evidence Act requires agencies to submit a capacity assessment, which is to provide a comprehensive view of evaluation and research capacity. The capacity assessment is expected to provide a baseline against which agencies can measure improvements of the effectiveness, quality and methods of evaluation and research. The aim is to improve agency’s ability to co-ordinate and increase technical expertise for evaluations, consequently increasing quality and use. The Evidence Act has also introduced Learning Agendas

to reinforce the learning aspect within departments and agencies. These are evidence-building plans to be included in the agency strategic plan, which identify and address priority questions relevant for programmes and policies, with strategies, data and methods to develop evidence. Finally, agencies conduct annual Strategic Objective Reviews to identify areas where additional evaluation, analyses, skills or other capacity are needed, and to develop a culture focused on learning and performance.

1.3.5. Role of the Parliament and Supreme Audit Institution

Governments are not able to bind the actions of their successors. For this reason, external institutional actors are important to support the evaluation policy of governments. Both Parliament and the Supreme Audit Office have a role to play to nurture and maintain a good evaluation system.

Parliaments can have an impact on the evaluation system in different ways.

Parliament as initiator of evaluations (in co-operation with Supreme Audit Office)

Parliaments sometimes initiate or produce evaluations. In many countries, parliaments co-operate with supreme audit institutions. Parliament also sometimes commissions evaluations to government, for instance by introducing evaluation clauses into laws.

- In Germany, the Bundestag (the lower House of Parliament) requires approximately 80 reports annually from the Federal Government regarding the evaluation of policies and measures of administrative action of the government. In some cases, Parliament organises evaluations, often through hearings resulting in parliamentary resolutions.
- In the United States, the Government Accountability Office (GAO) produces recommendations to Congress focused on how to improve the efficiency, effectiveness and responsibility of government operations and to use taxpayers' money appropriately. The majority of GAO's work consists of performance audits; it prepares some 600 – 700 reports each year and makes the information available publicly on its website. GAO's audits are initiated upon request by public congressional committees, on the independent initiative of the Comptroller General, and/or as prescribed by law. Also at the service of the Congress, the Congressional Research Service provides objective and confidential legislative analysis of current policies and the impact of proposed policy alternatives.
- In Canada, the Parliamentary Budget Office (PBO)'s mandate is to provide independent analysis to Parliament on the budget, as well as matters of particular significance relating to the nation's finances or economy; and at the request of a committee or a parliamentarian, to estimate the financial cost of any proposal that relates to matters over which Parliament has jurisdiction. By providing independent and non-partisan financial and economic analysis, the PBO supports Parliament with the goal of raising the quality of parliamentary debate on questions of public money and promoting greater budget transparency and accountability. However, the Canadian PBO has no specific mandate in terms of evaluation.

Parliament as user and demander of policy evaluations

Parliaments are major users of policy evaluations and have been instrumental in increasing evaluation use by promoting the importance of evaluative evidence in the budgetary cycle and by requesting more performance data on government spending. This is done through public hearings and instruments for parliamentary scrutiny of the executive. Moreover, the direct institutional relationship in many countries with Supreme Audit Institutions (SAI) allows the Parliament to receive the SAI's self-initiated evaluations in addition to performance audits, to be used in the political debate. The Parliamentary Office Bodies often facilitate the work of the Parliament by commenting in a structured way evaluations that are produced by the ministries/departments or by scrutinising the information delivered by the government (OECD, 2020[1]). Parliaments can be informed of evaluation activities and involved in scrutinising evaluation results:

- In Canada, the parliamentary committees may use the evaluation reports, now called Departmental Results Reports – which are accessible on the internet - to accomplish their mission of making government accountable for its activities. The Parliament is also the recipient of all programme evaluations produced by departments as well as the departmental results reports.
- In Germany, the Parliament's interest in evaluation has grown over time: committees frequently ask for scientific reports and can request impact assessments of draft bills. The Bundestag can also decide autonomously to introduce the obligation for the federal government to carry out an *ex post* evaluation and present the findings to Parliament. Similar to what happens in the United Kingdom, Members of the German Parliament can question the ministers' and the government's work, asking information on projects' performance and evaluations.
- In the Netherlands, the legislation stipulates that the government must inform the Parliament on both planned and executed evaluations and spending reviews, and has to provide performance monitoring information that allows the Parliament to exercise the control over the Ministers. A policy review is always sent to the House of Representatives, not only when finalised, but also in its design phase to validate the concept. Law proposals put forward by government for deliberation in parliament have to be accompanied by an explanation of the objectives, efficiency and effectiveness of the proposal.
- In the United Kingdom, the Parliament has an important role in the field of evaluations, as it has oversight of the operations of government through its various parliamentary committees. In particular, through the Public Accounts Committee, the Parliament uses the National Audit Office's Value for Money reports to call on government officials to answer questions regarding the implementation of policies and programmes and the efficiency of public spending. These sessions are publicly broadcasted and receive media coverage. Other commissions of parliament can also use the NAO reports for accountability purposes.
- In the United States, over time, Congress has requested an increasing volume of evaluation and performance information. Departments have responded by making more information available online, to prevent material to Congress becoming overwhelming and to make it available in a timely manner. Congress can then use performance information when taking decisions on authorising or re-authorising federal programmes.

Parliament as gatekeeper for a qualitative evaluation system

As Parliaments contribute to ensure accountability, they can play an important role in promoting a structured or systematic approach to conducting evaluations. Some Parliaments demand governments to submit a report, each year, on the policy evaluations carried out and how the results of these evaluations were reflected in policy. These reports may be either standalone documents or integrated in the budget documents. In the Netherlands for example, the parliament has been an important driver for the government's efforts to continuously improve the evaluation framework. In particular, the recent budget framework reform, the introduction of the *ex ante* evaluation, and the operation Insight in Quality (IIK) to improve the quality of the evaluation system, were all instigated by the Parliament. Similarly, the comprehensive framework for evaluations and evidence-informed performance information has been built mostly through the legislation enacted by Congress in the United States.

Over the last decade, the Supreme Audit Institutions of several countries have also conducted a systematic audit of the executive's evaluation system (Germany, United Kingdom and United States), which have identified the main gaps or weaknesses and triggered improvements.

An innovative example of how Parliament can contribute to quality and use of evaluations, is a pilot initiative started by the Parliament in the Netherlands to co-operate with academic researchers. At the request of the Parliament (Lower House), academics support the members of Parliament in assessing the way article 3.1 of the Accountability Act is applied to new policy proposals. The pilot was conducted for three legislative proposals. For each, academics were asked to assess whether the information provided by the government according to article 3.1 of the Accountability Act was sufficient to allow the Parliament to assess whether the policy goals were sound and measurable and whether there is enough room for learning and improvement, based on a well-designed mid-term evaluation (Tweede Kamer, 2019[15]).

How to facilitate an increased role by Parliament

Despite the fact that the Parliament can play an important role in this field, as some country examples show, in general the evidence from the five countries in the sample suggests that discussion of evaluation findings in Parliament generates less interest than could be expected.

The countries in this study have found several ways to increase the interest and impact of Parliament:

- In Canada, the Parliamentary Budget Office (PBO) does not have a specific mandate for policy evaluation. Nevertheless, it provides independent and non-partisan financial and economic analysis, which supports Parliament in raising the quality of parliamentary debates on questions about the use of public funds and promoting greater budget transparency and accountability.
- In the Netherlands, the government must inform the Parliament on both the planned and executed evaluations and spending reviews, and has to provide performance monitoring information that allows the Parliament to exercise control over the ministers.
- In the United Kingdom, the Scrutiny Unit is part of the Committee Office in the House of Commons and provides support and advice to select committees to enable them to better interpret, analyse and scrutinise financial information delivered or published by the government.
- In the United States, the Congress is supported by the Congressional Budget Office (CBO) and the Congressional Research Service.

Another response can be to provide training and expertise for elected representatives on policy evaluation, or the instalment of a specific committee for policy evaluation oversight, such as the Public Accounts Committee in the United Kingdom. This provides a platform to discuss work on policy evaluation, following the theme of evaluation and the planning that was agreed upon.

1.3.6. What interactions between central and subnational governments on policy evaluation?

In the benchmark countries, it seems that there is not much formalised interaction between the central and subnational governments on the subject of policy evaluation. In most cases, subnational governments define when and how to carry out evaluation. Each subnational government is responsible for its own approach and prioritisation of the use of evidence, policy evaluation and performance information. Even when the federal government transfers funds to the provinces or regions, which are then responsible for the implementation of the policies, there are generally some reporting or evaluation requirements to the central government or oversight. For instance, in Canada, transfer payments are based on respect for the accountability mechanism of each order of government to citizens.

In several countries, national government uses soft tools to interact with subnational governments instead of formal co-ordination mechanisms. For example, in Canada the federal-provincial relationship is a negotiated model that relies on co-operation. The Federal government organises policy tables with the provinces to discuss the delivery issues that fall within provincial responsibility. Issues pertaining to evaluations are not discussed at the policy table but separately as part of a professional practice. Specific frameworks for discussing evaluations are well developed in some areas, for example concerning labour market issues. In the specific area of the labour market transfer agreements, federal/provincial/territorial evaluation committees have been established to jointly develop and carry out evaluation activities. These committees represent a venue where networks of evaluators can share best practices, common knowledge and evaluation results. These committees ensure that evaluations are of high quality and satisfy the requirements set out in the bilateral agreements. As part of the bilateral agreements, information and data sharing arrangements also enable the conduct of state of the art impact analysis. Both federal and provincial actors benefit from this synergy. The goal is to enable a dialogue based on expertise. This model to evaluation represents a good example of different levels of government working collaboratively to inform improvements in programmes for the citizens.

In the Netherlands, since 2006, every municipality is required to install either a local policy evaluation function or commission (*rekenkamerfunctie*) or an independent local audit office (*rekenkamer*). The implementation is left to municipalities. Municipalities may also set up a joint audit office with one or several other municipalities. These local audit offices determine their own agenda and carry out investigations into the effectiveness, effectiveness and legality of local policies. This research can be *ex post* and/or *ex ante*. Often, an external research partner conducts the evaluations. The research reports (including the findings, conclusions and recommendations contained therein) are made public. Reviews of the system showed that in many municipalities, thorough research is absent and the importance of evaluations is not recognised. Small municipalities with smaller budgets to spend on research regularly use light methods (quick scans, roundtables with citizens, etc.) that can also be useful. The central government's subsequent attempts to encourage municipalities to make improvements have had little effect. The Dutch government therefore decided in 2019 to require an independent audit office and thus to diminish the autonomy of municipalities in this field. In addition, some investigatory powers were extended and the possibility of councillors acting as advisory members of an audit office was added to increase the involvement of the city council in the audit office. More investment is also being made in training for court researchers, and in improving co-operation and knowledge sharing between court auditors. The government is also promoting the creation of a common audit office for smaller municipalities to pool their budgets, support, knowledge and expertise.

Sometimes, for example in the United Kingdom and the Netherlands, there is some level of co-ordination between the National Audit Office (NAO) and the audit offices of the devolved administrations, which allows them to exchange good practices and keep regular communication, but the NAO cannot carry out value for money evaluations of programmes and policies implemented in the devolved administrations.

Co-operation is more developed or is being developed on the subject of collection of data and access to data, reflecting a will or need to gather ‘whole of government’ data, from Federal as well as subnational governments (United States, Netherlands). But overall, data sharing between central and regional governments is a challenge, as shown again by the COVID-19 crisis.

Co-operation is also present in the form of consultations on federal evaluations (priorities), where requirements of agencies to consult stakeholders include States or local governments (e.g. United States, on the preparation of Learning Agendas).

1.4. The evaluation process

The evaluation process is an important element of an evaluation system. Among other mechanisms, guidelines play an important role in providing answers to the questions on how to conduct evaluations, who conducts evaluations, how to involve stakeholders and how to create transparency (OECD, 2020[1]).

1.4.1. How to conduct policy evaluations?

Evaluation guidelines

Guidelines, toolkits and other supporting documents on evaluation can assist policy makers in conducting policy evaluation successfully. All the benchmark countries have (or are developing) guidelines to support the implementation of policy evaluation across government. Such guidelines generally aim to better plan, commission and manage evaluations.

Evaluation guidelines usually focus on the identification and design of evaluation approaches, the design of data collection methods, stakeholder engagement in the evaluation process, the quality standards and independence of evaluations, reporting of evaluation results, and use of evaluation evidence.

- Canada has published a significant number of guidelines for the implementation and evaluation of policies. These guidelines are designed to support programme managers and evaluation managers in activities on the development of performance measurement (Guide to Developing Performance Measurements Strategies (Government of Canada, 2010[16]) or the Rapid Impact Evaluation (Government of Canada, 2017[17]), but can also be more conceptual (Theory-Based Approaches to Evaluation: Concepts and Practices (Government of Canada, 2017[18]) or Integrating Gender-Based Analysis Plus into Evaluation: A Primer (Treasury Board Canada, 2019[19]).
- Germany does not currently have general guidelines at the federal level on how to conduct *ex post* evaluations or monitor the effectiveness of programmes, but they are being developed. Paragraph 7 of the General administrative regulation for the Federal Budget Code (VV-BHO) contains some elements of guidance, but these are relatively limited. Some ministries have published internal guidance documents. The National Regulatory Control Council (NKR), as an independent advisory body, has developed an evaluation model for an evaluation procedure which can be used as a possible basis for evaluation standards by the federal ministries (NKR, 2017[20]).
- In the Netherlands, the guidelines on Policy Reviews (Ministerie van Financiën, n.d.[21]) present the different steps to follow for conducting a policy review. These guidelines are complemented with additional information on thematic aspects and background information (Ministerie van Financiën[21]). In 2021, the Ministry of Finance also made available a toolbox[3] which provides information on methods and techniques for evaluation (Rijksoverheid[22]).
- In the United Kingdom, the Magenta Book provides guidance on policy evaluation: its scoping, design, management, use and dissemination (UK HM Treasury, 2020[23]). The Green Book is rather focused on the *ex ante* appraisal of expenditure and presents an overview on what expenditure appraisal and evaluation is, and how it fits within the government decision-making processes (UK HM Treasury, 2020[24]). It also provides guidance on the design and use of monitoring and evaluation before, during and after the implementation of policy interventions, and on the presentation of results. Both guidebooks are complemented with other manuals. These manuals are not mechanical or deterministic decision-making devices: they provide thinking models and methods to support officials conducting appraisals and evaluation. These documents also provide information on practical matters such as public procurement procedures for selecting evaluators or the use of evaluation results by policy makers.

- In the United States, the guidelines are mainly focused on how to organise the evaluation activities. The OMB publishes guidance material on the various phases of the implementation of the Evidence Act, and thus on Learning Agendas, programme evaluation, etc.

Guidelines have to simultaneously respect the desired methodological consistency and allow for the necessary tailor-made approach within each individual evaluation. They have to be systematically updated to integrate experience and new developments. The Green Book in the United Kingdom for instance has been updated regularly since its inception (the last update was end 2020).

Evaluation standards

Evaluation standards on the quality of the evaluation process can be part of the evaluation guidelines, but can also exist separately. Evaluation standards set out more general ethical principles applying to evaluation activities but can also include methodological guidance. Evaluation standards can also be produced by evaluation professionals themselves, targeting not only officials conducting evaluations but all evaluation activities.

- The Canadian Evaluation Society (CES) has adopted programme evaluation standards and uses them to give trainings. Five standards for a good evaluation plan are identified: utility, feasibility, propriety, accuracy and accountability, and thus are not an ethical guide.
- The German Evaluation Society has also established similar standards (utility, feasibility, propriety and accountability).
- The American Evaluation Association has published the Guiding Principles for Evaluators (AEA, 2018[25]). These standards reflect the core values of the Association and are intended as a guide to describe the professional ethical conduct of evaluations.

The standards set by professional organisation can be complemented with standards established by the government. In the United States, the OMB published the Program Evaluation Standards and Practices (OMB, 2020[26]) to guide agencies in developing and implementing evaluation activities and in hiring and retaining qualified staff. These standards are related to relevance, utility, rigor, independence and objectivity, transparency and ethics. In Canada, the Treasury Board maintains a Standard on Evaluation (Directive on Results, Appendix C) that requires departments to consider factors such as risk, relevance, effectiveness, efficiency, and government-wide policy considerations (e.g., gender-based analysis, official languages) when conducting evaluations.

1.4.2. Who conducts the policy evaluations?

Internal or external evaluations?

As described in the previous chapter, line ministries are responsible for conducting evaluations in the selected countries. However, this does not necessarily imply that officials within the ministries undertake all evaluations themselves. An important choice to be made is about “who” carries out an evaluation: external or internal evaluators? (OECD, 2020[1])

- **External** evaluations are conducted outside of the institution in charge of the public policy, that is, either by another government institution or by an institution outside of the government. This type of evaluation could be considered more independent, it can look more critically at the policy being studied and its results can be potentially more trusted. However, it can be limited by the knowledge of the evaluator about the context and political process, as well as access to relevant data.
- **Internal** evaluation refers to an evaluation conducted by the institution in charge of the public policy that is being evaluated. Internal evaluators may have more knowledge about public policies, provide a more accurate assessment and have easier access to inside data, but they can also

have mixed incentives when it comes to assessing and reporting on how well a policy has performed. In the process of conducting an evaluation, internal evaluators can be under political pressure and time constraints, which can affect the validity of the findings of the evaluation and its public deliberation.

The decision in the benchmark countries about whether to do policy analysis in house or externally involves a range of considerations including: the skills and expertise available in-house, the scale of the project, the timeframe for completion and any budgetary constraints. There is no one-size-fits-all solution and there is a range of risks and issues with conducting work in house and contracting it out.

- In Canada, evaluation activities are both internally conducted and outsourced to external experts. The choice between the two is *ad hoc* and is decided by the line ministry (due to complexity, limited internal capacity, etc.).
- In Germany, there is a common sense and usual practice that evaluations are outsourced to external stakeholders in order to ensure objectivity. Usually, these comprise scientists from universities and research institutes.
- In the Netherlands, the majority of evaluations are conducted by external parties, partly because of the lack of internal capacity in most line ministries and partly to increase their objectivity.
- In the United States, some agencies conduct evaluations in-house while others rely on independent external evaluators, and some do both. Agencies generally outsource evaluation activities depending on the subject matter and the capacity needed to conduct a specific analysis, as well as whether the agency desires an objective external assessment.

In practice, differences between internal and external evaluations are often blurred because of hybrid approaches, mixing internal and external evaluations for specific or more technical aspects of a given policy. Moreover, a government can commission the evaluation to an external organisation, while still ensuring that civil servants control the research questions addressed by the evaluation, or help with the gathering of data. Similarly, an internal evaluation can be accompanied by oversight or a steering group of ‘outside’ stakeholders and experts (OECD, 2020[1]).

In the Netherlands, the guidelines on Policy Reviews recommend to include an independent evaluator in the process (Ministerie van Financiën[21]). There are several possibilities to include this independent input in the evaluation exercise, for instance by involving this person in advisory meetings or in the actual evaluation exercise. It is also mandatory that, after the evaluation has been concluded, an independent opinion of the policy review is made and sent to the House of Representatives. This can be done by the same person who was involved in the evaluation or by another person.

An important issue is to have in-house capability for evaluations, both to conduct them internally as well as to oversee those commissioned externally. Consultants can usefully supplement the expertise available within government, but how they may best contribute to specific cases needs careful consideration, and they should not be over utilised to the detriment of internal capability.

Just as for in-house evaluations, it is important to guarantee the quality of outsourced evaluations. In some countries, every policy analysis project that is commissioned must go through the specialised unit for checks and approval. An example of this can be found in the US Department of Justice, which observed that the use of third parties to conduct evaluations or review performance does not necessarily ensure independence or neutrality. As such, the integrity and control processes for evaluations in the department apply to outsourced evaluations as well as evaluations that are completed using in-house resources.

Finally, it should be mentioned that in some countries advisory councils, independent research institutes and universities play a significant role as suppliers of policy evaluations, often at their own initiative.

Capacity building

Capacity building and embedding learning in the organisational culture is crucial. The knowledge, skills and abilities required to conduct policy evaluation have to be available at the right moment and in the right place. In the benchmark countries a variety of practices are implemented, focusing on capacities and skills.

A first question about capacity is where it needs to be developed, namely at the centre of government level but also within the line ministries, where this can be done within a dedicated unit or throughout different entities within the ministry.

Based on the case studies, it appears that all countries with the exception of Germany put a strong emphasis on the availability of capacity at the centre of government. This allows the central level to prepare central guiding and to be involved in the evaluation work done by line ministries (as a peer reviewer or in a quality control role).

- In Canada, there are 50 analysts working in the TBS Results Unit and they are encouraged to participate in training as well as on-the-job learning. The TBS focuses on capacity building both at the central level of the departments as at the lower level. Within line ministries, the Policy on Results gives to TBS the responsibility to establish, amend or rescind competencies for heads of performance measurement, heads of evaluation and evaluation specialists across departments, guaranteeing the capacity at the top of the departments and making them responsible for the capacity within the departments. It is important to note that departments can also develop very significant evaluation programmes, with for example over 70 staff at employment Canada above. So, the total number of staffs and resources engaged in this agenda at the federal level can be very significant.
- The Netherlands has a similar system. Each ministry has to establish a directorate for Financial and Economic Matters (FEZ) to supervise evaluations of the effectivity and efficiency of policies, among other tasks. The way the FEZ are organised is not prescribed, nor does the Ministry of Finance provide guidelines for it. In general, these units are relatively limited in size and capacity and the way the FEZ are involved in capacity building activities towards the officials within their ministries depends on their own initiatives. In some ministries, there is a separate unit responsible for evaluations, while in others programme managers are responsible for evaluations.
- In the United States, the Evidence Act requires agencies to carry out a capacity assessment led by the Evaluation Officer. The assessment is intended to improve agencies' ability to co-ordinate and increase technical expertise for evaluations, thus focusing on capacity building throughout departments and agencies. Within the Department of Health and Human Services for instance there is a specific unit responsible for organising seminars and training sessions to enhance the evaluation capacity within the department.
- Germany is the exception, in the sense that there is no obligation to create specific units in line ministries for the evaluation function, leaving it in the hands of the line ministries to organise themselves and take care of the capacity building.

Evaluation profession

In Canada and the United States, legislation mandates specific functions linked to the evaluation process and describes the responsibilities attached to this position. The Netherlands and United Kingdom have a different approach.

- In Canada, the Treasury Board Secretariat has published a document listing the competencies for the Heads of Evaluation. Moreover, the Canadian Evaluation Society has been committed, since 2009, to the implementation of the title of accredited evaluator. This professional title programme is built on three pillars: a code of ethics, professional standards and a set of 36 theoretical and practical skills (knowledge of basic concepts of assessment, mastery of data collection and processing methods, project management skills, interpersonal skills, etc.).

- Although in the Netherlands there is no specific legislation for the evaluation profession, some departments have appointed Chief Science Officers. This is a scientist who provides support and direction to (evaluation) research, and can also build a bridge between academic researchers and the ministry, aiming at more evidence-informed policy development and policy-relevant research.
- Although this is not a common practice, in the United Kingdom, evaluation work is mainly done by two specific professions specialised in evaluation activities within the civil service: the government economic service and the government social research profession. For these, profession standards have been elaborated (such as the Policy Profession Standards) describing the skills and knowledge required by policy professionals at all stages of their career and provide a framework for professional development. Moreover, NESTA with the Alliance for Useful Evidence has created ‘Evidence Masterclasses’, providing “an immersive learning experience” for senior decision makers who want to become more skilled and confident users of research.
- In the United States, the Evidence Act requires that federal departments designate an Evaluation Officer, Chief Data Office and a Statistical Official, each with well-described tasks.

1.4.3. How to create transparency and involve stakeholders?

Transparency of the evaluation system, the involvement of stakeholders throughout the evaluation process and public accessibility of evaluations are key elements to ensure quality, gain support to establish and maintain a sound evaluation system and strengthen trust in government and its delivery of evidence-informed policies.

Tools and practices to promote transparency

- **Publicly available evaluation agendas** are a standard procedure in the Netherlands. The agenda for the planned Policy Reviews is transmitted to the Parliament. For the newly created Strategic Evaluation Agenda, a similar procedure will be developed, allowing the Parliament and the public to know the plans in advance, in order to comment on them. In addition, the Netherlands has also a repository of all planned Policy Reviews that can be consulted on the web. In the US, the Program Evaluation Standards and Practices stipulates that Federal evaluations must be transparent in their planning, implementation and reporting phase, and the Foundations for Evidence-Based Policymaking Act includes a provision that requires agencies to submit and publish annual evaluation plans. In Canada, approved evaluation reports and summaries, including complete management responses and actions plans, are released on web platforms.
- **Publishing evaluation results**. This is a common practice in the Netherlands. In Canada, departments have to make all their evaluations public, whether or not they are produced internally or externally. In the United Kingdom, the Magenta Book on evaluations states that evaluation findings and materials should be public by default. In practice, the publication of evaluations depends on the level of importance. In Germany, evaluations that are conducted by external actors are published, but internal evaluations are not necessarily so.
- **Presenting the results of evaluations to the Parliament as well as how the results will be used**. Government feedback is included in the Policy Reviews that are presented to the Parliament in the Netherlands.
- **Developing communication strategies and using tools such as social media**. In Canada, departments are diffusing evaluation findings beyond departmental websites via such platforms as Twitter and LinkedIn. Evaluations are also published on the federal government website.

- **Creating databases for evaluations accessible on the internet.** In the five benchmark countries there is no evaluation database that gives an overview of all the evaluations that have been conducted, contrary to information on performance monitoring where several countries have portal sites containing references to this information in a consolidated matter. In the Netherlands, the Ministry of Finance keeps track of all Policy Reviews on a website, but there is no database containing evaluations.
- **Keeping track of the implementation of evaluations' recommendations**. Periodic publication of a report on the status of policy evaluation and how the results have been reflected in the policy-making process can be a stimulus to improve the quality. In November 2017, a report commissioned by HM Treasury and led by Sir Michel Barber, "Delivering better outcomes for citizens: practical steps for unlocking public value" ("Public Value Report"), triggered a renewed attention on outcomes and results (Barber, 2017[27]). In the Netherlands, the Operation Insight in Quality that has been conducted the past years was the result of an external report on the quality of budget expenditures.

1.5. Conditions for robust evaluation systems

The governance of policy evaluation as well as the process, including the aspects of capacity, clearly have an impact on the robustness of a policy evaluation system. In this section the focus will be on another aspect, namely how the benchmark countries try to guarantee the quality of policy evaluations (fostering the necessary skills as well as providing peer reviews and control mechanisms), cope with the difficulty of data availability and how they promote the use of evaluations.

1.5.1. How to ensure quality?

The quality of evaluations is essential to ensure impact on policy making, and thus that evaluations actually serve as tools for learning, accountability and better decision making. Bad quality evaluations can lead to poor evidence, which can be costly and misleading, and threatens political support to invest resources for policy evaluation (OECD, 2020[1]). Both legislation and capacity building in all their aspects aim to improve the quality of the evaluations, but this may not be enough.

A variety of mechanisms for quality control of evaluations and quality assurance can be put in place to foster evaluations that are technically and methodologically sound and well-governed (OECD, 2020[1]). These mechanisms can be organised from the main actors in charge of evaluation across government or created at the level of individual ministries/departments.

Internal co-ordination and learning mechanisms

Some countries have developed co-ordination bodies or mechanisms to allow evaluators to share good practices.

- In Canada, co-ordination and feedback mechanisms between branches of government take place through both formal and informal structures. The Results Division has recently created two steering committees, one for evaluation and the other for performance measurement. The goal is to discuss challenges and develop concrete actions to address them. There are regular meetings (about 2-3 per year) where all the Heads of Evaluation and Heads of Performance Measurement are invited. With the creation of the new steering committees, the approach is evolving. There is a sense that working with a smaller group of Heads on specific issues and then validating the approach with the broader group, will be more effective in working collaboratively on specific files and implementing concrete actions (e.g., promoting the value of the functions, improving data quality and availability, sharing best practices, recruitment of staff, professional development, etc.).

- In the Netherlands, within the Ministry of Finance, the Bureau for Strategic Analysis (BSA) has created an evaluation community (Interdepartmental Supervisory Committee on Performance and Evaluation or IBP), mainly consisting of people working in FEZ. They meet regularly (monthly or bimonthly) to exchange good practices and discuss relevant approaches for evaluations and policy reviews.
- In the United Kingdom, there are also several bodies which support and facilitate the development of guidance and the sharing of practices related to policy evaluation. For instance, the Government Economic and Social Research Team is a professional unit based in HM Treasury that provides support across government departments, co-ordinates learning and development activities, develops profession standards, and provides mechanisms for developing standards for evaluation of policy and social research. This team also co-ordinates the Cross-Government Evaluation Group, the network of evaluators where good practices and common challenges are shared. The Group is cross-departmental and cross-disciplinary, made up of analysts and evaluation managers from government bodies, and works to improve the supply of, stimulate the demand for, and encourage the use of good quality evaluation evidence in government decision making.
- In the United States, an interagency council has been set up, bringing together Evaluation Officers and serving as a forum for exchanging information and advising OMB on issues affecting the evaluation functions, such as the evaluator competencies, best practices for programme evaluation, and evaluation capacity building.

In-house training

In-house training by government is common, but can also be provided by the departments or the Ministries themselves. This can be included as an obligation in the evaluation framework, or conducted at their own initiatives.

- In Canada, policy evaluation capacity is developed and improved mainly through training. The Results Division organises information and learning sessions (called Drop-in Sessions) based on different results-related subjects of interest to the evaluation and performance measurement communities, and has recently started to record and post these sessions on an unlisted Results Division YouTube channel for their communities to view. Presentations of other departments' and non-governmental organisations' workshops are also shared via the Results Portal when relevant to the evaluation and performance measurement communities. Guidelines are not just uploaded on the website: the TBS delivers presentations, newsletters, etc., to communicate to people and disseminate knowledge. The line departments, particularly those with the largest capacity for evaluation, also develop evaluation capacity through workshops, learning series and individual training tailored to evaluators needs.
- In the United Kingdom, the central government supports capacity development of its evaluators in various ways, such as through the establishment and support of a network of evaluators to share good practices and common challenges (Cross-Government Evaluation Group), and advisory panels for evaluations (Cross-Government Trial Advice Panel). Moreover, government departments also carry out internal training programmes, such as the Ministry of Business, Innovation and Skills which annually offers technical training on scientific assessment methods.
- In the United States, the OMB established an inter-departmental council of evaluation officers in 2019 to support capacity building on evaluation as a cross-governmental function. Other capacity-building activities undertaken by OMB include technical assistance through workshops and seminars to support the implementation of its guidance memoranda, as well as resources and other materials posted on central websites. At the agency level, the Department of Health and Human Services has also organised seminars and training sessions to explain the purposes of the Evidence Act and how it applies to the department.

Evaluation societies

Evaluation societies sometimes offer training courses and workshops on evaluation and thus play a significant role in some countries in this respect.

- The Canadian Evaluation Society (CES) brings together nearly 2000 practitioners from different backgrounds (administrations, universities, or private consulting firms). It promotes leadership, knowledge, advocacy, and professional development and offers dedicated training. The training, which is not compulsory and is separate from the accreditation procedure, offers awareness-raising actions at the margins of civil servants who are not members of ministerial assessment units.
- The German Evaluation Society (DeGEval) is the main entity focused on the professionalisation of evaluators and the development of a theoretical framework. It has published the Standards for Evaluation, the Guidelines for implementation of the Standards in the Field of Self-Evaluation, the Recommendations on Education and Training in Evaluation (which define the requirement profiles of evaluators) and the Recommendations for Clients of Evaluation. DeGEval has also established working groups to serve as a platform for specific dialogue in various fields of application and are involved in annual conferences and meetings, which ministries also regularly participate in.
- Also in the United States, evaluation is recognised as a profession and is represented and supported by the American Evaluation Association.

Peer Review

During an evaluation, public consultation, advisory panels and committees can be used to promote the quality of evaluations. Their main aim is to provide comments and feedback throughout the different phases of the evaluation (design, data collection, synthesis, etc.). These advisory panels and committees may be composed of policy practitioners, evaluations experts and stakeholders, and may be established on an ad hoc basis or systematically.

Peer review is a mechanism that is used by several countries to guarantee and improve the quality of evaluations. In Canada, the Treasury Board Secretariat encourages peer review. These also represent a widespread practice in Germany, particularly in the field of development co-operation.

In the Netherlands, it is recommended to include an independent evaluator in the process. There are several possibilities to include this independent input in the evaluation exercise, for instance through an advisory body or by involving this person in the actual evaluation process. It is also mandatory that, after the evaluation exercise has been concluded, an independent opinion of the policy review is provided and sent to the House of Representatives. This can be done by the same person who was involved in the evaluation or by another person.

Control and oversight mechanisms

An important way to guarantee and improve quality is to set-up of a control mechanism. Control mechanisms are rare, and exist mainly for *ex ante* regulatory impact assessments.

In the countries under study, there is no strong centralised control on the quality of the (*ex post*) evaluations done within line ministries. Some countries have organised an internal control mechanism at the level of the line ministries. External oversight by, for example, the Supreme Audit Office is also important as a quality control mechanism.

- In Canada, there is a control entity in larger departments. According to the Policy on Results, the Performance Measurement and Evaluation Committee within the larger departments reviews evaluation reports. There is no institutionalised control mechanism at the central level (TBS).

- In Germany, the Federal Court of Auditors (Bundesrechnungshof) follows a comprehensive auditing approach: its work cover aspects of financial compliance and performance audits, either separately or in combination. Performance audits study the criteria of economy, effectiveness and efficiency to ensure that value for money is obtained according to Section 7 of the Federal Budget Code. Despite not having an evaluation mandate, the Federal Court of Auditors may be seen as assuring quality when performing its audits, as it also checks the adherence to existing evaluation requirements. Moreover, the 2019 decision to update the concept of *ex post* evaluation establishes that each evaluation report will be subject to a quality control before publication, by an entity to be chosen by ministries (BPA, 2019[12]).
- In the Netherlands, there is no obligation to create an evaluation committee such as in Canada. Some ministries have created voluntarily an evaluation committee, although there are no standard guidelines for this, and every minister is free to organise it according to its own insights. Typically, this committee consists of senior officials responsible for a directorate, external expert(s) and a delegation of the FEZ. This committee can discuss the planning and implementation of the evaluation activities in the ministry, and can also discuss some major evaluation exercises.

1.5.2. Data and information availability?

The quality and availability of non-evaluation specific data (big data, open data, statistical data, administrative data, etc.) in a format that can be readily used is also a factor that influences how easily a policy can be evaluated.

A good evaluation system relies on comprehensive, quality data. The full implementation of an evidence-based agenda implies leveraging the data that are available for analytical purposes. As such, a high quality national statistics system is an integral part of any evaluation strategy, as well as up-to-date databases and registers that mutually communicate and disaggregate data at the desired level.

Important data sources for policy evaluation are (OECD, 2020[1]):

- Statistical data: commonly used in research, it corresponds to census data or more generally to information on a given population collected through national or international surveys.
- Administrative data: this data is generally collected through administrative systems managed by government departments or ministries, and usually concerns whole sets of individuals, communities and businesses that are concerned by a particular policy. For instance, it includes housing data and tax records.
- Big data: mainly drawn from a variety of sources such as citizen inputs and the private sector, big data is most often digital and continuously generated. It has the advantage of coming in greater volume and variety.
- Evaluation data: this data is collected for the purpose of the evaluation. It can take the form of qualitative questionnaires, on-site observations, focus groups, or experimental data. See further down for a description of impact evaluation methods to collect and analyse data.

Often, the limited availability and quality of data across government agencies and departments is a major challenge for evaluation practices. In many countries, there is no integrated data infrastructure and access to data across departments remains difficult. The architecture often does not guarantee data interoperability. Data siloes remain within the government itself, as departments operate in similar sectors but cannot (or refuse to) share data. Problems remain with obtaining linked files where there is a low coverage of common identifiers to merge datasets. In some cases, the challenge is to understand what data and data sets currently exist within departments, and how departments could use the data for policy analysis.

Apart from technical problems, enabling the strategic use and quality of data requires human capabilities, especially the willingness of public entities and civil servants to share use data, and investment in databases, data linkages and analytical tools. Political reluctance towards sharing evidence on policy impact and effectiveness may also be another barrier in accessing data.

In several cases, for instance in co-operation between national and subnational governments, data sharing and availability has proven to be a topic for improvement, notably with regard to access to information and data. Access to data is complicated as they are gathered by different entities and there is no obligation to share data with the federal level or vice versa. Often, subnational Governments enjoy full autonomy in some specific matters and this issue has become as a challenge during the COVID-19 pandemic (for example, in Canada and Germany).

In recognition of the limitations in infrastructure to support the use of evidence, some jurisdictions have launched initiatives to try to maximise the use of government's existing assets for evidence-informed policy making. These include measures to avoid fragmentation and duplication of efforts (e.g. in developing separate data sharing infrastructures) and to promote public sector integration and cohesion, library facilities, data portals and clearinghouses as well as data sharing software such as the use of APIs and open data and other methods of maximising government's data assets.

Fostering a Data Driven Public Sector culture can be a very effective way to enhance the quality of ongoing evaluations through the application of relevant data (OECD, 2019[28]). Data-Driven Public Sector initiatives generate an environment in which data about policy interventions is available in real time to avoid waiting for monthly or quarterly updates across a wide range of policy areas because the data they need is more frequently available and accessible.

- In Canada, the Data Strategy Roadmap for the Public Service is intended to support a more strategic use of data while protecting citizens' privacy. While not all data have privacy implications, when they do, departments and agencies should incorporate privacy by design and engage early with the Office of the Privacy Commissioner (OPC). Co-location of TBS and Statistics Canada staff was a success factor with combined teams jointly working on validation of results. Another success factor was the engagement of several actors involved in evaluations, i.e. departments and agencies, as well as the research community. The Canadian government has also worked to support the "demand side" of open data and information, identifying and collaborating with stakeholders in organisations and companies that leverage open government data and information.
- In the United States, the Foundations for Evidence-Based Policymaking Act was designed in part to ensure that the necessary data quality and review structures were in place to support the use of administrative data in evaluations. The Evidence Act incorporates the Open Government Data Act, which requires agencies to publish information online as open data, using standardised and machine-readable data formats. The Act emphasises co-ordination to advance agencies' data management and data access functions by mandating an open government approach to data. The website Data.gov, launched in 2009 and managed by the US General Services Administration (GSA), provides access to government datasets on a wide range of topics. The GSA has to maintain a 'Federal Data Catalogue' as an online point of entry dedicated to sharing agency data assets with the public. Guidance is being prepared by OMB for Open data access and management and for Data access for statistical purposes. Agencies are also required to designate a Statistical Official, to advise on statistical policy, techniques and procedures. The Statistical Official collaborates and consults regularly with the Chief Data Officer (CDO) and Evaluation Officer. Finally, agencies develop and maintain a comprehensive data inventory that accounts for all data assets created and collected by the agency. The OMB has established an Advisory Committee on Data for Evidence Building at the federal level, to review, analyse and make recommendations on how to promote the use of federal data for evidence building and how to facilitate data sharing and data linkage.

However, challenges remain, including finding and accessing the right data, having common data definitions and terminologies, overcoming privacy-related issues, and mobilising resources to make the kind of longer-term investments required to develop new data linkages of good quality.

- In Germany, the Federal Statistical Office is the leading provider of statistical information required for developing informed opinions and decision-making processes. The Federal Statistical Office conducts the nationwide statistics in close co-operation with the statistical offices of the *Länder*.
- In the Netherlands, Policy Reviews should be a synthesis of existing information, but it appears that in most policy areas there is not enough evaluation information available to base the policy review solely on these previous evaluations. Therefore, the questions have been updated recently to include an improvement paragraph, that answers the following question: are there gaps in the available information to make a relevant evaluation of the effectiveness and efficiency of the policy and how can these gaps be filled?
- The What Works Centres in the United Kingdom are institutions that provide independent, evidence-based, and practical advice in their policy area. Most of the What Works Centres were founded in the 2010s, although some of these centres existed before. In 2013, the Cabinet Office and HM Treasury launched the What Works Initiative with the objective of improving the supply of high-quality evidence, creating incentives to use evidence in policy decisions and practice, and building capability across government to improve the use of evidence for policy making. The What Works Initiative established the What Works Network, a network where all What Works Centres are gathered. All What Works Centres are engaged to generate evidence on what works in a defined policy area, translate evidence for specific user groups in a friendly-format, and encourage the adoption and use of evidence in decision making. To integrate the What Works Network, a What Works Centre should fulfil some criteria: it must commit to the principles of the Network, share learning across the Network, engage with national and local exercises, and participate on the decisions regarding prospective new members joining the Network (Cabinet Office, 2018[29]).

Finally, in order to have the relevant data for *ex post* evaluations, it is important to think about it from the start of a policy.

- Germany has recently introduced a provision along this line. The 2013 concept for evaluation was updated in 2019, when a decision of Secretaries of State clarified additional aspects to be included in the draft bills that would facilitate and improve quality of *ex post* evaluations. When presenting a draft bill, its objectives and the criteria for achieving the objectives have to be stated, as well as the steps and methods of evaluation that will be used. Furthermore, the Federal government is developing a guidance document outlining the steps and methods of an evaluation.
- This issue has also clearly been stated in the Netherlands, where in 2018 the Accountability act was amended to make sure that when preparing new policies, the goals have to be established clearly form the start as well as an indication how the proposal will be evaluated after implementation (evaluation paragraph). This is one of the outcomes of the discussion as part of the operation Insight in Quality, on the link between *ex ante* and *ex post* evaluation. An important role of the FEZ is to stimulate the programme officers to start thinking about the policy reviews far ahead of the time at which the policy review has to be conducted, making sure that the necessary data are available when the policy review starts.
- In the United Kingdom, policy proposals are required to envisage monitoring activities and indicators, so that corrective actions can be taken during policy implementation. For instance, single departmental plans include indicators that can be used to monitor performance on each objective. After policy implementation, government departments are expected to conduct *ex post* evaluations, including project implementation review and post-evaluation review. *Ex ante* and *ex post* evaluations should be linked, as the policy design should identify the relevant indicators to be monitored and evaluated, and ensure that the information is collected during the policy process.

1.5.3. How to promote use of evaluations?

A sound framework and correct implementation of evaluations is no guarantee for successful use of evaluations in the policy process. Therefore, uptake of policy evaluations results should be enhanced. A strong demand from Parliament and external stakeholders is a key driver for an increased use of evaluation information, although the combination of a well working evaluation framework and the incentives to use the resulting evaluation information is still no guarantee for an optimal use.

Stakeholder engagement is an important way to ensure that evaluation improves policy making. First, and most obviously, stakeholders (namely ministries and agencies but also the public) can improve the quality of evaluations by providing information on the impacts of a policy, as well as providing feedback on preliminary analysis and findings from evaluations. Second, engaging stakeholders early in the evaluation process can help identify which policy areas need evaluation the most. Third, giving the public the opportunity to express views and make an input during evaluation procedures can build trust in the evaluation process and even a sense of 'ownership' of the outcomes, making the implementation of any changes politically easier to manage (OECD, 2020[1]).

In Canada, stakeholders are not consulted in the planning stages of evaluations, but can be engaged in the implementation phase, as part of participatory processes. According to the Directive on Results, the head of evaluations has to consider the information needs of the major stakeholders when identifying the department's five-year evaluation coverage needs. In the Netherlands, the newly created Public Value Scan requires ministries to consider civil society, by indicating the value added for society and the support for the policy within target groups and stakeholders, therefore encouraging them to interact with civil society on these questions. In the US, OMB guidance mandates stakeholder involvement in developing learning agendas and planning for evaluation within agencies. Moreover, one of the tasks of the Chief Data Officer Council, composed of each agency Chief Data Officer, is to consult with the public and engage with stakeholders on how to improve access to data assets, which are relevant for evaluations and evidence-informed activities.

Some other mechanisms to promote the use of evaluations include:

- Conducting user-focused or utilisation-focused evaluative processes. The Public Value framework in the United Kingdom is very focused on optimising the process of turning funding into policy outcomes for citizens. The newly introduced Public Value Scan in the Netherlands requests each minister to indicate what value the pursued policies add to society when evaluating a policy.
- In order to match evidence with demand of policy makers and users' needs in terms of timing and priorities, evaluation agendas have recently been introduced in the Netherlands (Strategical Evaluation Agenda) and in the United States (Learning Agenda).
- The Evaluation Task Force that was created in the United Kingdom in 2020 aims to ensure that evaluation and evidence are an integral part of spending and operational decisions, by incorporating them into HM Treasury spending processes, and supporting government departments to generate evidence in priority areas with evidence gaps.
- Evidence synthesis, using standardised formats, meta-analysis and systematic reviews, etc. In Germany, the IZA World of Labour, an academic-supported online platform, publishes a comprehensive set of understandable, non-technical summaries of policy relevant research on labour market themes. The summaries ('one-pager') always have the same structure and there is a link to a more extensive article of 10 pages.
- Embedding use in the institutional set-up, within and outside of the executive:
    - The policy on Results in Canada supports the inclusion of evaluation and evidence in the budget process through the Treasury Board Submissions (a key financial decision stage for the Cabinet). Furthermore, the Treasury Board is working with the Privy Council Office to

identify how tools developed within the PCO for a more outcome-focused approach to some policies can be leveraged by the TBS as conditions for the funding of departments.

- In the Netherlands, the policy reviews are sent to the House of Representatives with a cabinet response indicating how the governments intends to act on the results of the evaluations. The fact that the central steering of the evaluation framework in the Netherlands is embedded in the Ministry of Finance, implies that the evaluation framework is linked with the budget process.
- The policy profession standards require British civil servants who belong to the policy profession (i.e. the civil servants who are involved in the design of public policies) to use scientific data in their work. This means that they are required to base their public policy recommendations on evaluation results.

- Promoting communication, availability and accessibility of evaluations. Transparency and public availability of data and evaluation results can stimulate the use of evaluation findings and encourage ministries and stakeholders to make use of the results, as in the United States where Congress has requested an increasing volume of evaluation and performance information.
- Organising interaction between analysts and policy makers. In Germany, the Evaluation Society (DeGEval) brings together academics and officials and has also established working groups to serve as a platform for specific dialogue in various fields, bridging both worlds.

Moving from knowledge management to knowledge brokerage, acknowledging the fact that the worlds of policy making and research are very different. In several countries there are specific institutions within the Parliament that are tasked to translate technical evaluation reports into more user-friendly reports that focus on the needs of the Parliaments in order to increase the use of evaluation in Parliamentary work. The Analysis and Research Department in the Netherlands and the Scrutiny Unit in the United Kingdom are two examples.

References

AEA (2018), *Guiding Principles For Evaluators*, American Evaluation Association, Washington D.C., https://www.eval.org/About/Guiding-Principles#:~:text=The%20five%20Principles%20address%20systematic,how%20they%20justify%20professional%20actions. [25]

Barber, M. (2017), *Delivering better outcomes for citizens: practical steps for unlocking public value*, https://www.gov.uk/government/publications/delivering-better-outcomes-for-citizens-practical-steps-for-unlocking-public-value (accessed on 13 December 2021). [27]

BPA (2019), *Decision of the State Secretaries Committee on Bureaucracy Reduction of 26 November 2019*, Press and Information Office of the Federal Government, Presse- un Informationsamt der Bundesregierung, https://www.bundesregierung.de/resource/blob/974430/1698788/784303d11758802127d37fc38f49dc8a/2019-11-27-beschluss-evaluierung-data.pdf?download=1. [12]

Cabinet Office (2018), *What Works Network Membership criteria*, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/747965/WW-membership-paper_Oct2018.pdf. [29]

Government of Canada (2017), *Guide to Rapid Impact Evaluation*, Centre of Excellence for Evaluation (CEE), https://www.canada.ca/en/treasury-board-secretariat/services/audit-evaluation/centre-excellence-evaluation/guide-rapid-impact-evaluation.html. [17]

Government of Canada (2017), *Theory-Based Approaches to Evaluation: Concepts and Practices*, Centre of Excellence for Evaluation (CEE), https://www.canada.ca/en/treasury-board-secretariat/services/audit-evaluation/centre-excellence-evaluation/theory-based-approaches-evaluation-concepts-practices.html. [18]

Government of Canada (2010), *Supporting Effective Evaluations: A Guide to Developing Performance Measurement Strategies*, Centre of Excellence for Evaluation (CEE), https://www.canada.ca/en/treasury-board-secretariat/services/audit-evaluation/centre-excellence-evaluation/guide-developing-performance-measurement-strategies.html. [16]

Government of Canada (1985), *Financial Administration Act*, https://laws-lois.justice.gc.ca/eng/acts/F-11/FullText.html (accessed on 13 December 2021). [10]

International Organisation of Supreme Audit Institutions (2019), “ISSAI 100: Fundamental Principles of Publc Sector Auditing”, https://www.issai.org/wp-content/uploads/2019/08/ISSAI-100-Fundamental-Principles-of-Public-Sector-Auditing.pdf (accessed on 26 June 2021). [9]

Ministerie van Financiën (2021), *Strategische Evaluatie Agenda*, https://www.rijksfinancien.nl/memorie-van-toelichting/2022/OWB/IX/onderdeel/1075313 (accessed on 13 October 2021). [14]

Ministerie van Financiën (n.d.), *Guidelines on Policy Reviews*, https://archief.rijksfinancien.nl/handreiking-beleidsdoorlichtingen (accessed on 13 December 2021). [21]

NKR (2018), *Germany: Less Bureaucracy, More Digital Services, Better Regulation. Let's get to it! - Annual Report 2018*, National Regulatory Control Council, https://www.normenkontrollrat.bund.de/resource/blob/656764/1548226/a53ca395512296087f96a45d0b839d44/2018-11-09-jahresbericht-englisch-data.pdf?download=1. [4]

NKR (2017), *Annual Report. Bureaucracy Reduction. Better Regulation. Digital Transformation. Leverage Successes – Address Shortcomings*, https://www.normenkontrollrat.bund.de/resource/blob/656764/771786/137189edf779b7cf4e7c564a60763756/2017-annual-report-data.pdf?download=1 (accessed on 13 October 2021). [20]

OECD (2021), *Monitoring and Evaluating the Strategic Plan of Nuevo León 2015-2030: Using Evidence to Achieve Sustainable Development*, OECD Public Governance Reviews, OECD Publishing, Paris, https://doi.org/10.1787/8ba79961-en. [8]

OECD (2020), *Improving Governance with Policy Evaluation: Lessons From Country Experiences*, OECD Public Governance Reviews, OECD Publishing, Paris, https://doi.org/10.1787/89b1577d-en. [1]

OECD (2019), *The Path to Becoming a Data-Driven Public Sector*, OECD Publishin, Paris, https://doi.org/10.1787/059814a7-en. [28]

OECD (forthcoming), "OECD best practices for spending reviews". [7]

OECD (forthcoming), "Recommendation of the Council on Public Policy Evaluation", *OECD Legal Instruments*, OECD, Paris. [2]

OMB (2020), *Circular No. A-11*, Executive Office of The President, Office of Management And Budget, Washington, D.C., https://www.whitehouse.gov/wp-content/uploads/2018/06/a11.pdf. [6]

OMB (2020), *Phase 4 Implementation of the Foundations for Evidence-Based Policymaking Act of 2018: Program Evaluation Standards and Practices*, Executive Office of The President, Office of Management And Budget, Washington, D.C., https://www.whitehouse.gov/wp-content/uploads/2020/03/M-20-12.pdf. [26]

Rijksoverheid (n.d.), *Policy Evaluation Toolbox*, https://www.toolboxbeleidsevaluaties.nl/ (accessed on 13 December 2021). [22]

State Secretaries Committee on Bureaucracy Reduction (2013), *Concept for the evaluation of new regulatory proposals*, National Regulatory Control Council, https://www.normenkontrollrat.bund.de/resource/blob/822080/1565438/0da55de5ab47b1a6986eff0212f4e6d7/evaluation-data.pdf. [13]

Treasury Board Canada (2019), *Integrating Gender-Based Analysis Plus into Evaluation: A Primer*, http://www.canada.ca/en/treasury-board-secretariat/services/audit-evaluation/evaluation-government-canada/gba-primer.html (accessed on 13 October 2021). [19]

Treasury Board Canada (2016), *Policy on Results*, https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=31300. [3]

Tweede Kamer (2019), *Tweede Kamer, Brief van de Vaste Commissie voor Financiën voor de verbetering verantwoording en begroting (Vergaderjaar 2019-2020, nr. 176)*, http://www.tweedekamer.nl. [15]

UK HM Treasury (2020), *The Green Book: appraisal and evaluation in central government*, https://www.gov.uk/government/publications/the-green-book-appraisal-and-evaluation-in-central-governent. [24]

UK HM Treasury (2020), *The Magenta Book*, https://www.gov.uk/government/publications/the-magenta-book. [23]

UK HM Treasury (2018), *Guidance The Green Book: appraisal and evaluation in central government*, https://www.gov.uk/government/publications/the-green-book-appraisal-and-evaluation-in-central-governent. [5]

US Congress (2019), *Foundations for Evidence-Based Policymaking Act of 2018*, https://www.congress.gov/bill/115th-congress/house-bill/4174/text (accessed on 13 December 2021). [11]

Notes

[1] The five case studies and the synthesis note were produced as part of a project requested by the Independent Authority for Fiscal Responsibility of Spain (AIReF) and funded by the European Commission (DG Reform), to improve the quality of public expenditure and policy making in Spain.

[2] Policy reviews provide a synthesis of individual evaluations that have been carried out over the previous years within a policy area, examining the effectiveness and efficiency of the policies. They are mandated by law at least once every 7 years.

[3] https://www.toolboxbeleidsevaluaties.nl/

From:
OECD Journal on Budgeting

Access the journal at:
https://doi.org/10.1787/16812336

Please cite this article as:

Mathot, Axel and Flavia Giannini (2022), “Evaluation Framework and Practices: A comparative analysis of five OECD countries”, *OECD Journal on Budgeting*, Vol. 22/2.

DOI: https://doi.org/10.1787/911cc792-en