



OECD Public Governance Reviews

Improving Governance with Policy Evaluation

LESSONS FROM COUNTRY EXPERIENCES



OECD Public Governance Reviews

Improving Governance with Policy Evaluation

LESSONS FROM COUNTRY EXPERIENCES

This document, as well as any data and map included herein, are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.

The statistical data for Israel are supplied by and under the responsibility of the relevant Israeli authorities. The use of such data by the OECD is without prejudice to the status of the Golan Heights, East Jerusalem and Israeli settlements in the West Bank under the terms of international law.

Please cite this publication as:

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>.

ISBN 978-92-64-68184-2 (print)

ISBN 978-92-64-82982-4 (pdf)

OECD Public Governance Reviews

ISSN 2219-0406 (print)

ISSN 2219-0414 (online)

Photo credits: Cover Image created by OECD GOV.

Corrigenda to publications may be found on line at: www.oecd.org/about/publishing/corrigenda.htm.

© OECD 2020

The use of this work, whether digital or print, is governed by the Terms and Conditions to be found at <http://www.oecd.org/termsandconditions>.

Preface

In a world marked by fragmentation, growing citizen demand and multiple tipping points, governments are confronted with increasingly complex, interrelated challenges. The COVID-19 pandemic has presented such challenges on an unprecedented scale, obliging governments to quickly develop and roll out bold policy responses and extensive expenditure packages.

Especially in times of crisis, maintaining citizens' trust requires decision makers to be accountable for how well policies work and that they work for all. Failing to rise to the policy evaluation challenge may entail serious consequences for economies and societies, as governments confront the need to make significant progress in evidence-informed policymaking.

The report *Improving Governance with Policy Evaluation* outlines how OECD countries have developed systemic approaches to policy evaluation. It presents experiences and lessons learned from investing in institutionalisation and promoting quality of evaluation over the past decade. The report provides a unique overview of the institutional frameworks that enable policy evaluation, and promote its quality and use. The many practices, concrete experiences and examples it contains aim at facilitating knowledge sharing and strengthening policy implementation in OECD countries and beyond.

It is essential to equip policy makers with the necessary tools to identify effective solutions for smart, responsive and agile government – this report can help them get there.

A handwritten signature in black ink, appearing to read 'Jeffrey Schlagenhauf', with a large, stylized initial 'J'.

Jeffrey Schlagenhauf

OECD Deputy Secretary-General

Foreword

Policy evaluation is a critical element of good public governance. Policy evaluation can help ensure public sector effectiveness and improve the quality, responsiveness and efficiency of public services. Ex ante evaluation feeds into the policy-making process at the design and implementation phase, informing, for instance, the design of new rules or the allocation of resources. Evaluation is also essential ex post, to draw lessons and to provide an understanding of what works, why, for whom, and under what circumstances. Policy evaluation connects policies, policy makers and citizens, helping ensure that decisions are rooted in trustworthy evidence and deliver desired outcomes.

This report offers a new, cross-cutting contribution to the global policy debate on evaluation and evidence-informed policy making. The need for a thorough understanding of evaluation emerged from OECD's efforts to define a holistic approach to sound public governance by taking stock of the lessons from all the governance policy communities. There were no systematic comparative studies on policy evaluation systems and cultures across OECD countries. Moreover, there was a need to link up the various elements that relate, directly or indirectly, to policy evaluation, including regulatory practices, performance budgeting, and supreme audit institutions.

This report offers a comprehensive analysis of the institutionalisation, quality and use of evaluation from a systemic perspective. This implies an analysis of each of these dimensions and how they are related to each other to ensure evaluation contributes decisively to sound public governance. The report relied on a survey of 42 OECD and non OECD countries, which is the first significant cross-country survey of policy evaluation practices in an OECD context. The report presents the results of this survey together with examples of good practices from countries. It also draws on results from other data on performance budgeting, centres of government and regulatory policy.

Generally, countries show a strong commitment to policy evaluation, as this is embedded in a range of legal and policy frameworks and even at the level of the constitution for some countries. However, implementing policy evaluation remains a challenge for many, and this reflects an unfinished policy agenda. This report sheds new light both on the challenges and the policy responses that are developed across countries. These seek to mobilise a range of tools and to invest in skills and organisations to promote the use and quality of evaluation.

Overall, the report seeks to foster knowledge-sharing in an area that remains in many ways a frontier. It offers evidence to guide countries seeking to implement evidence-informed policy-making strategies and to improve public sector effectiveness. The report can be a useful tool for strengthening the capacity for policy implementation and for learning. Finally, sharing and promoting good practices in this area is also important for improving citizens' trust in governments' decision making processes and to enable sound public governance in a complex and fast changing social and economic environment.

This study was carried out under the auspices of the OECD Public Governance Committee, which approved the document for publication on 10 June 2020.

Acknowledgements

The report was prepared by the OECD Directorate for Public Governance (GOV) under the leadership of Marcos Bonturi, as part of the work programme of the Governance Reviews and Partnerships Division, headed by Martin Forst.

The report was prepared under the strategic direction of Stephane Jacobzone, Head of Unit for Evidence, Monitoring and Policy Evaluation. Iván Stola, policy analyst, and Claire Salama, junior policy analyst, co-ordinated the report. Emma Cantera, policy analyst, and Lizeth Fúquene, drafted the first chapter. Iván Stola and Simon Schmitz, junior policy analyst, drafted the second chapter on the institutional underpinnings. Claire Salama drafted the third chapter on promoting quality and use with inputs from Anne-Pauline de Cler. General support was provided by Diego Millan, Shashwat Koirala, Jaylia Yan and Naoki Usuda.

The report was prepared following a data collection exercise, with a survey initiated by Eva Beuselinck, Head of Unit at the time, together with Iván Stola and Simon Schmitz. The survey also benefited from valuable comments from Daniel Acquah, Rebecca Schultz, as well as inputs from Alessandro Lupi, Zsuzsanna Lonti, Daniel Gerson and Santiago Gonzalez from the Public Governance Directorate. Claire Salama and Simon Schmitz conducted the quality assurance process of the data. The team would like to thank officials from Belgium and Mexico who helped to pilot the survey. The authors are also grateful to all the country representatives who participated in the survey and who provided valuable feedback on the data collected through the quality assurance process.

This report benefitted from insights and comments on the conceptual and methodological framework, and on the analysis from many experts, colleagues and country officials. In particular, the authors would like to thank all the participants and experts who took part in an expert meeting in September 2019 which reviewed preliminary results. The report also benefitted from the roundtable organised in the context of the Senior Budget Officials Network on Performance and Results, including Aline Pennisi and Alessandra de Lellis, Ministry of Finance Italy, Christine Minas from the Treasury Board Secretariat Canada. The authors are also grateful to Stephan Naundorff, Federal Chancellery Germany, Chair of the Regulatory Policy Committee, and to Diana Epstein, US Office of Management and Budget for their participation in the Public Governance Committee's discussion in November 2019. At the OECD, the report benefitted from valuable insights from Christiane Arndt, Pauline Bertrand, Andrew Blazey, Monica Brezzi, Julio Bacio Terracino, Mariana Prats, Bagrat Tunyan, Barbara Ubaldi, Gregor Virant, Benjamin Welby in the Public Governance Directorate.

Table of contents

Preface	3
Foreword	4
Acknowledgements	5
Executive Summary	10
Chapter 1. Towards a systemic approach to policy evaluation	12
Key findings	13
Introduction	13
Why does policy evaluation matter?	13
Government core objectives for policy evaluation	14
Defining policy evaluation	15
Towards sound policy evaluation systems: promoting institutionalisation, quality and use	27
Note	29
Chapter 2. Making Policy Evaluation happen: What are the institutional underpinnings?	34
Key Findings	35
Understanding the institutional set-up for policy evaluation	35
Anchoring policy evaluation in legal & policy frameworks	38
The principal institutions in charge of policy evaluation and their mandates	50
Notes	66
References	66
Annex 2.A. Annex guidelines and methods	80
Chapter 3. How do countries address the challenges of promoting quality and use of evaluations?	83
Key findings	84
Introduction	84
Quality and use are essential	85
Exogenous factors affecting quality and use of evaluations	86
Promoting quality through good governance and sound methodology	89
Ensuring evaluation for impact: promoting the use by politicians, practitioners and citizens	120
Notes	152
References	152

Annex A. Data and methodology

166

Survey Responses	166
Process for data collection and validation	167

Tables

Table 1.1. Comparing policy monitoring and policy evaluation	16
Table 1.2. Type of evaluations	25
Table 2.1. Nature of legal framework for policy evaluation	39
Table 2.2. Features of the framework for policy evaluation	46
Table 2.3. Mandate of Centre of Government for policy evaluation	54
Table 2.4. Guidelines and methods for policy evaluation	80
Table 3.1. Mechanisms for the promotion of quality	91
Table 3.2. Quality standards included in evaluation guidelines	93
Table 3.3. Impact evaluation methods	98
Table 3.4. Methods used by countries in the evaluation of government-wide policy priorities	100
Table 3.5. National Evaluation Societies in OECD countries	119
Table 3.6. Mechanisms to promote the use of evaluations	122
Table 3.7. Mandates of coordination platforms	143
Table A.1. ISO codes of the countries referred to in this report	167

Figures

Figure 1.1. Government's main objectives for conducting evaluations	15
Figure 1.2. Government's formal definition of policy evaluation	18
Figure 1.3. Conceptual clusters included in the definition of policy evaluation	19
Figure 1.4. Actors involved on carrying out evaluations of government-wide policy priorities (GWPP)	23
Figure 1.5. What policies are being evaluated?	24
Figure 1.6. Policy evaluation criteria along the policy chain	25
Figure 1.7. Elements in the policy cycle chain that are evaluated	26
Figure 1.8. Challenges for promoting policy evaluation across government	27
Figure 2.1. Availability of a legal framework guiding policy evaluation across government	38
Figure 2.2. Availability of a policy document on policy evaluation across government	45
Figure 2.3. Institutions within the Executive that have competences related to policy evaluation across government	51
Figure 2.4. Governance of <i>ex ante</i> and <i>ex post</i> evaluation	53
Figure 2.5. Mandate of the Ministry of Finance/Economy/Treasury or equivalent	58
Figure 2.6. Mandate of the Ministry of Planning, Development, or equivalent	60
Figure 2.7. Institutions beyond the executive that have competences for policy evaluation at central/federal level	62
Figure 3.1. Countries that have standards for quality evaluations in their policy/legal framework	104
Figure 3.2. Peer reviews	106
Figure 3.3. Systematic and meta-evaluations	107
Figure 3.4. Competence requirements for evaluators	110
Figure 3.5. Training for internal or external evaluators	111
Figure 3.6. A specific job category in government	112
Figure 3.7. Advisory panels/ steering committees	114
Figure 3.8. Establishment of a network of evaluators	115
Figure 3.9. Audit of the policy evaluation system by Supreme Audit Institutions	116
Figure 3.10. Requirements related to stakeholder engagement in policy/legal frameworks	123
Figure 3.11. Types of stakeholders engaged in policy evaluations	124
Figure 3.12. Stakeholder engagement in the evaluative process	125
Figure 3.13. Publicity of evaluation results	128
Figure 3.14. Guidelines containing standards for the use of policy evaluation	133
Figure 3.15. Management response mechanisms at the level of specific institutions	139
Figure 3.16. Coordination platforms across government to promote use of evidence	142

Figure 3.17. Incorporation of policy evaluation findings into the budget cycle	145
Figure 3.18. Influence of evaluation findings on budget allocation decisions	145
Figure 3.19. Formal requirements to conduct RIA and <i>ex post</i> evaluation of primary laws	147
Figure 3.20. Discussion of evaluation findings at the Centre of government	148
Figure 3.21. Discussion of evaluation findings in Parliament	151

Boxes

Box 1.1. Definitions of (policy) evaluation	16
Box 1.2. Policy evaluation criteria	19
Box 1.3. Public interventions: Costa Rica	20
Box 1.4. Quality attributes in countries definitions	21
Box 1.5. Components of a sound policy evaluation system	29
Box 2.1. Examples of policy evaluation-related principles found in national constitutions	40
Box 2.2. France's embedded policy evaluation framework	41
Box 2.3. Building the institutional foundations for evidence-based policymaking in the US	42
Box 2.4. Institutionalisation of Policy Evaluation in Japan	43
Box 2.5. Korea's Policy Evaluation System	44
Box 2.6. Canada's Policy on Results	47
Box 2.7. The National Evaluation Policy (PNE) in Costa Rica	48
Box 2.8. The role of frameworks and guidelines for the promotion of evidence-informed policy making (EIPM) in Canada	49
Box 2.9. Basic Guidelines for Implementing Policy Evaluation by the Ministry of Internal Affairs and Communications (MIC) of Japan	50
Box 2.10. Examples of institutions responsible for Policy Evaluation in OECD countries	52
Box 2.11. The role of the Ministry of Finance in institutionalising policy evaluation: the experience of the Netherlands'	56
Box 2.12. Key practices for value-for-money assessment- a key responsibility for the Ministry of Finance or equivalent	57
Box 2.13. Committee for Monitoring and Evaluation of Federal Public Policies (CMAP) in Brazil & Colombia's National Planning Department	59
Box 2.14. The National Council for the Evaluation of Social Development Policy of Mexico (CONEVAL)	61
Box 2.15. The role of autonomous agencies in Italy	61
Box 2.16. Chile's Supreme Audit Institution's role in strengthening good governance	63
Box 2.17. The French Supreme Audit Institution's role in conducting policy evaluation	63
Box 2.18. A cross-country perspective at Supreme Audit Institutions' role in policy evaluation	64
Box 3.1. Benefits of developing an intervention theory of change and logic model for policy or programme development	87
Box 3.2. Potential sources of data used for policy evaluation	88
Box 3.3. The example of the Magenta Book in the United Kingdom: Core questions of policy evaluations	94
Box 3.4. Standards for commissioning evaluations in Costa Rica	95
Box 3.5. Korea's office for government policy Coordination (2017) framework act on government performance evaluation	96
Box 3.6. Data collection standards in guidelines	97
Box 3.7. Evaluation guidelines in Great Britain	99
Box 3.8. Understanding independence in evaluations	101
Box 3.9. Australia's productivity commission: An autonomous government body	102
Box 3.10. Sources of conflict of interest in evaluations	103
Box 3.11. The review function of Japanese Ministry of Internal Affairs and Communications (MIC)	105
Box 3.12. Internal and external peer reviews in Portugal and Germany	106
Box 3.13. Meta-evaluations	108
Box 3.14. Self-evaluation checklists in Spain and Poland	109
Box 3.15. Training Evaluators in Slovakia	112
Box 3.16. Policy evaluation as a profession in Ireland and the UK	113
Box 3.17. Audits of the evaluation system by SAs: Examples from Belgium and Estonia	117
Box 3.18. The role of the European Court of Auditors in auditing the evaluation system of a Directorate General	118
Box 3.19. The experience of Alberta in Canada and the "What We Heard" report	126
Box 3.20. The concept of Utilization-Focused Evaluations	127

Box 3.21. Evaluation portals to promote the use of evidence	129
Box 3.22. Different methodologies for reviewing the evidence base	131
Box 3.23. The United Kingdom's What Works Network	132
Box 3.24. Guidelines for the use of evaluation evidence	134
Box 3.25. Guidelines for the construction of Evidence Gaps Maps: A tool for decision making in Colombia	135
Box 3.26. The skillset for Evidence Informed Policy Making	136
Box 3.27. Intensive skills training and mentoring programmes	137
Box 3.28. Management response mechanisms in Mexico and Costa Rica	139
Box 3.29. The role of knowledge brokers	140
Box 3.30. The experience of Australia's Productivity Commission in communication and public inquiries	141
Box 3.31. The role of the US Office of Management and Budget in promoting the use of evaluation	142
Box 3.32. Use of evaluation findings in the budgetary cycle in Canada	146
Box 3.33. Discussion of evaluation results at the Council of Ministers in Korea	149
Box 3.34. The Brazilian Committee for Monitoring and Evaluation of Federal Public Policies	149

Executive Summary

Policy evaluation contributes fundamentally to sound public governance. It can help governments improve the design and implementation of public policies that can, in turn, lead to prosperity for their country and well-being for its citizens. Policy evaluation contributes to promoting public accountability, learning and increased public sector effectiveness through improved decision-making.

The report provides a broad analysis across 42 countries of the institutionalisation of policy evaluation, its quality and use. This systemic perspective allows for a full discussion of how evaluation can contribute to the policy cycle as well as the tools that rely upon evaluation, such as budgeting and regulation. A sound institutional set-up can help align isolated and unplanned evaluation efforts into more formal and systematic approaches. Promoting both better quality and use can have a greater impact in ensuring that evaluation fully achieves its purpose.

Generally, countries show strong commitment to policy evaluation. Some countries have embedded policy evaluations in their constitutions, and around two-thirds of responding countries have developed some kind of legal framework for policy evaluation. Similarly, most countries have adopted guidelines on policy evaluation applicable across government.

The term “evaluation” can cover a range of practices. Around two-thirds of the countries surveyed have at least one definition of evaluation. While these definitions of policy evaluation often reflect the specificities of a country’s institutional setting, common concepts can be found across definitions, in particular regarding what should be measured (policies, programmes, plans, reforms), why an evaluation should be conducted (aims), and when (ex-ante or ex-post).

Most countries face significant challenges in promoting policy evaluation across government. These challenges are mainly related to the limited use of evaluation results in policy making, the absence of a coherent whole-of-government strategy for policy evaluation, and the lack of human resources – whether in terms of skills, capacity or capability.

Sound institutional set-ups can provide incentives to ensure that evaluations are effectively conducted. They can promote transparency and accountability in the management of evaluations, and contribute to improving both the comparability and consistency of results as well as the use of results in policy processes. For this reason, countries are using a range of institutional approaches to anchor policy evaluations in their governance apparatus. They may also co-ordinate their approaches and frameworks for policy evaluation with those related to evidence and data governance.

In addition to legal frameworks, about half of the countries have also developed policy frameworks that provide principles and strategic direction for evaluations. Such policy frameworks provide high-level guidance and clarity for institutions by outlining overarching best practices and goals.

The centre of government provides strategic direction for policy evaluation in two-thirds of the countries surveyed. As such, the centre plays a crucial role in embedding a whole-of-government approach to policy evaluation and it often has the broadest mandate to do so. Ministries of finance also have significant responsibility in 60% of countries. These results show the strong economic impetus for policy

evaluation and the close connection between incentives to enhance the quality of public expenditure and incentives to deliver results. Ministries of public sector reform and planning also play a role in some countries, along with autonomous agencies. Agencies are often a source of good evaluation practices, particularly in the Nordic countries, even if their role remains decentralised. They are generally well placed to conduct independent, transparent and accountable policy evaluations.

Co-ordination mechanisms such as commissions and integrated cross-departmental services, when they exist, can help strengthen evaluation systems, for example in terms of staffing and capabilities. Locating the office in charge of policy evaluation close to political decision-making power may allow it to be more effective in commissioning policy evaluations and following up on commitments by ministries. At the same time, ensuring that evaluation systems are independent, transparent and accountable can help bolster citizens' trust in the results.

Policy evaluation can only be truly effective if it is of high quality and its results are used. Quality control and quality assurance are key to ensuring the robustness of evaluations. Standards play an important role in quality assurance, and guidelines exist in three-quarters of the countries surveyed. Other quality control mechanisms, such as peer review, systematic reviews, and competency requirements for evaluators are relatively common. Up to half of the countries organise training for evaluators, and a majority recognise the importance of developing evaluator competences.

While using the results of evaluations is a challenge, it is paramount to achieving impact. Countries are relying on a range of organisational and institutional mechanisms to promote their use and to create a marketplace for evaluations. Some of these measures consist of a co-ordination platform (in about one-third of the sample) or a management response (in one-quarter of the sample). Rating and grading systems are also used to a limited extent. Finally, most mechanisms to promote skills and competences are aimed at evaluators and managers, rather than to improve the capacity of policy makers and decision makers to use evidence.

The heterogeneity of country approaches suggests that the creation of an evaluation marketplace depends on the local political and cultural context. Evaluation can also be embedded into policy planning and policy-making processes. Half of the countries incorporate evaluation results in the budget cycle. In particular, many OECD countries use spending reviews. The area of regulatory policy is also one where the use of evaluation is well developed, with significant requirements for evaluation embedded in the regulatory impact assessment process.

The role of institutions outside of the executive branch remains limited, both in the promotion of quality and use of evaluations, beyond their involvement in the budgetary cycle – although Supreme Audit Institutions are a key actor in terms of the supply of evaluations.

Chapter 1. Towards a systemic approach to policy evaluation

Governments need to understand how and why a policy has the potential to succeed, and to ensure the efficient allocation of their financial resources. However, the understanding of the different practices used to assess whether government actions have met their expected goals and how they may complement each other, remains limited. This chapter provides an overview of policy evaluation across OECD countries and underlines the importance of developing a systemic approach in this area. The chapter discusses the relevance of policy evaluation and its distinctive role in the public sector and analyses countries' definitions of policy evaluation. The chapter also introduces the three components of policy evaluation systems: institutionalisation, quality and use of evaluation.

Key findings

- **Countries generally express strong commitment towards policy evaluation:** There is a shared concern to understand and improve government's performance and outputs, as well as to promote evidence-informed policy-making, and improve the quality of public services.
- **Policy evaluation is part of a range of practices geared to ensuring government's effectiveness and efficiency:** these include monitoring, spending reviews, and performance management. Not only do these practices complement each other, but policy evaluation also has a distinctive role to play in providing credible evidence for various public management efforts, such as monitoring or performance budgeting.
- **More than half of the countries (27 out of 41) have a formal definition of policy evaluation:** 14 of them have one definition applicable across the government, while in 13 several definitions coexist.
- **While countries' definitions on policy evaluation reflect their own institutional set up, common elements are present,** including what should be measured (policies, programs, plans, reforms), why an evaluation should be conducted (aims), when (ex-ante or ex-post), and the actors involved.
- **The most common criteria for evaluation are outputs and outcomes,** followed by policies processes and impacts.
- **Countries face several challenges for promoting policy evaluation across government** such as the limited use of evaluation results in policy-making, the absence of a strategy for policy evaluation that promotes a whole of government approach, the limited availability of human resources (capacities and capabilities) and the lack of an integrated approach to evidence management, including data.
- **A systemic approach, relying on mutually supportive elements in terms of institutionalisation, quality and use** is most likely to ensure a methodologically rigorous and systematic adoption of evaluations throughout the policy cycle, and use of findings by decision-makers.

Introduction

This chapter provides a first overview of the nature of policy evaluation across survey respondents and introduces the importance of developing sound policy evaluation systems. The first section discusses the relevance of policy evaluation for countries, outlines why policy evaluation matters, and addresses governments' main objectives for conducting evaluations. The second section analyses countries' definitions of policy evaluation, adopting an empirical approach. The last section aims to introduce this paper's approach to policy evaluation systems and its three components: institutionalisation, quality of evaluations and use of results.

Why does policy evaluation matter?

Governments are facing increasingly complex economic, social and environmental challenges, known as the VUCA, Volatile, Uncertain, Complex and Ambiguous, which require systemic approaches and system thinking. These challenges are compounded by increased citizen demand and higher expectations, fragmentation in knowledge, higher perception of corruption, skill gaps in the civil service budgetary constraints and an erosion of trust in public institutions (OECD, 2018^[1]). Lessons learned from OECD experience also highlight that it is more difficult for governments to identify outcomes, trade-

offs, as well as winners and losers of an implemented policy due to the increase complexity, divergent values and interdependent processes, structures and actors that are related to major policy challenges (OECD, 2018^[2]).

In this context, governments should demonstrate that their decisions and policies are informed by evidence, that they set realistic expectations about various policy choices, and spend public resources adequately. Thus, policy evaluation has a critical role to ensure these goals as well as to avoid policy failure (Howlett, 2019^[3]). By evaluating performance and results, policymakers have a deeper understanding of the underlying policy problems and can make informed decisions about the feasibility of continuing the policy or initiating a new one.

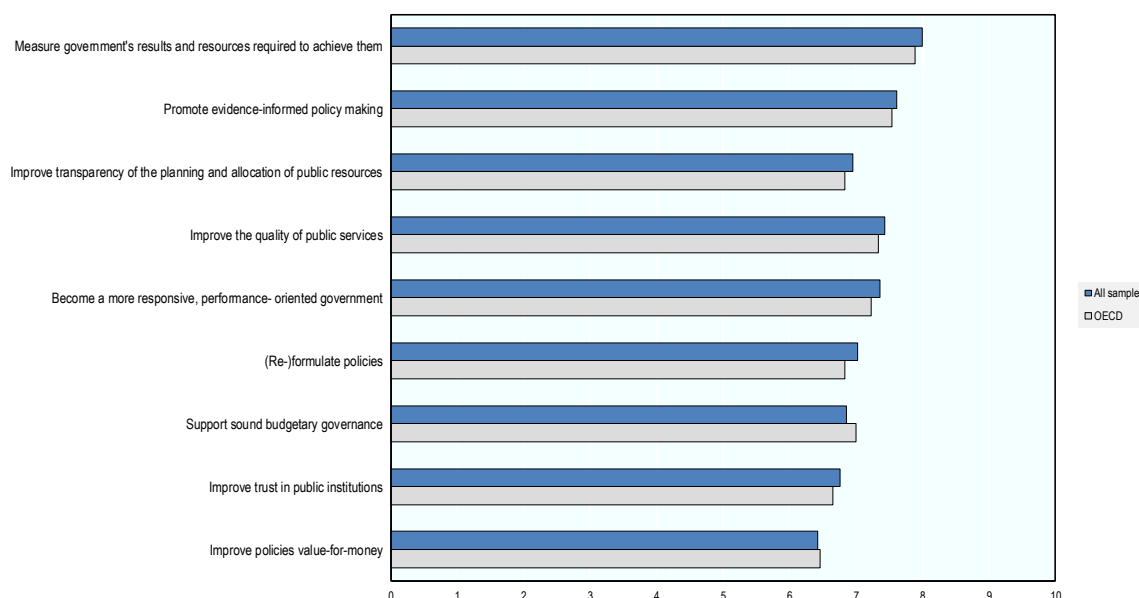
Policy evaluation facilitates **learning** as it helps to understand why and how a policy was or has the potential to be successful or not, by providing an assessment about the reasons and causal mechanisms leading to policy success or failure. It contributes moreover to the quality of decision-making by providing insights on how to improve links between policy formulation, implementation and outcomes (OECD, 2017^[4]). Simultaneously, policy evaluation has the potential to improve policy **accountability** and transparency, and provide legitimacy for the use of public funds and resources as it provides citizens and other stakeholders with information whether the efforts carried out by the government, including allocated financial resources, are producing the expected results (OECD, 2018^[2]).

Government core objectives for policy evaluation

The OECD survey on policy evaluation provides an overview of governments' stated objectives when conducting evaluations (See Figure 1.1). According to these results, countries ranked most of the objectives between 9 and 10, showing their strong commitment toward policy evaluation. The results show no clear-cut priorities among these objectives aside from their shared concern to measure government's performances/outputs and the resources required to achieve them, as well as to promote evidence-informed policy-making (OECD, 2018^[2]) and improve the quality of public services in all respondents and OECD countries. Countries are also concerned about conducting evaluations to improve policies value-for-money, to enhance trust in public institutions, and to encourage transparency in the allocation of public resources –albeit apparently to a slightly lesser extent.

Learning is also often an important objective of policy evaluation, even if it does not appear as such in the results below. It is often crucial and has been identified as such by lead experts (Lazaro, 2015^[5]). While the results in terms of learning are often less likely to achieve media impact or strong public attention, they are also potentially the most useful in that they can help improve policies and understand why policies work or don't, and what kind of adjustment may be needed.

Figure 1.1. Government's main objectives for conducting evaluations



Note: n=42 (35 OECD member countries). Answers reflect responses to the question, "What are the government's main objectives for conducting evaluations?", where 0 indicates "least important objective", 5 is "Neutral", and 10 is a "principal objective".

Source: OECD Survey on Policy Evaluation (2018).

Defining policy evaluation

What is the distinctive role of policy evaluation in the public sector?

For the purpose of this report, and as a reference for the survey respondents, *policy evaluation* is defined as a "structured and objective assessment of an ongoing or completed policy or reform initiative, its design, implementation and results. Its aim is to determine the relevance and fulfilment of objectives, efficiency, effectiveness, impact and sustainability as well as the worth or significance of a policy" ¹. While this definition may not be universally accepted by the evaluator community, it has offered a starting point to start the analysis and the questionnaire design. From an empirical perspective, a number of countries define evaluation, and others specifically do regarding policy evaluation. Box 1.1 presents some of these examples.

Box 1.1. Definitions of (policy) evaluation

The Netherlands: “Policy evaluation is an examination of the efficiency (the extent to which the optimum effect is achieved with as few costs as possible and undesirable side effects) and effectiveness (the extent to which the policy objective is realized through the use of the policy instruments examined) of policy.” (Ministry of Finance of The Netherlands, 2018^[6]).

United States: “Evaluation means an assessment using systematic data collection and analysis of one or more programs, policies, and organizations intended to assess their effectiveness and efficiency.” (115th Congress, 2019^[7]).

Canada: “Evaluation is the systematic and neutral collection and analysis of evidence to judge merit, worth or value. Evaluation informs decision-making, improvements, innovation and accountability. Evaluations typically focus on programs, policies and priorities and examine questions related to relevance, effectiveness and efficiency. Depending on user needs, however, evaluations can also examine other units, themes and issues including alternatives to existing interventions. Evaluations generally employ social science research methods” (Canada Treasury Board, 2016^[8]).

Source: 115th Congress of the United States (115th Congress, 2019^[7]), Ministry of Finance of The Netherlands (2018^[6]), Canada Treasury Board (Canada Treasury Board, 2016^[8]).

Promoting the comprehension of why and how a policy was or has the potential to succeed (i.e. learning) and improving the efficient allocation of financial resources (i.e. accountability) is becoming a priority across public administrations. However, there is a lack of awareness from practitioners and stakeholders about the different practices developed to assess whether government actions have met their expected goals (monitoring, spending reviews, or performance management), and how they differ from one another or support/complement each other. For example, while in some cases, policy evaluations can come close to performance audit, these two practices still differ in fundamental ways as professional disciplines.

This is why it is important to distinguish between monitoring and evaluation. Monitoring seeks to check progress against planned targets and can be defined as the formal reporting of evidence to show that resources are adequately spent, outputs are successfully delivered and milestones met (HM Treasury, 2011^[9]). (Table 1.1).

Table 1.1. Comparing policy monitoring and policy evaluation

Policy monitoring	Policy evaluation
Ongoing (leading to operational decision-making)	Episodic (leading to strategic decision-making). Differs from audit.
Monitoring systems are generally suitable for the broad issues/questions that were anticipated in the policy design	Issue-specific
Measures are developed and data are usually gathered through routinized processes	Measures are usually customized for each policy evaluation
Attribution is generally assumed	Attribution of observed outcomes is usually a key question
Because it is ongoing, resources are usually a part of the program or organisational infrastructure	Targeted resources are needed for each policy evaluation
The use of the information can evolve over time to reflect changing information needs and priorities	The intended purposes of a policy evaluation are usually negotiated upfront

Source: Adapted from McDavid, Huse and Hawthorn (2006^[10]), Program evaluation and performance measurement: an introduction to practice, Thousand Oaks, California: Sage, in OECD (2019^[11]), Open Government in Biscay.

Additionally, over the past years OECD governments have developed a range of tools that focus on strengthening the alignment of budget decision making with the government policy cycle, with a view towards improving performance and overall public sector effectiveness (OECD, 2019^[12]), such as:

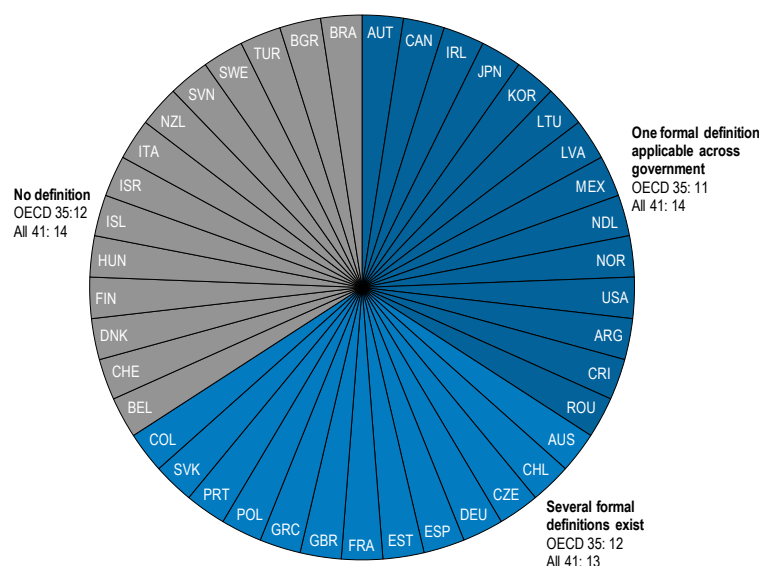
- **Spending reviews**, which aim to increase the fiscal space available to government to finance its policy priorities. Initially, after the global financial crisis, spending reviews were identified as a tool for extracting savings from agencies in a way that would seek to improve the quality of public expenditure. (OECD, 2011^[13]). Since then, reviews have evolved to identify whether a line agency's activities align to a government's priorities and the implementation challenges that an agency faces (OECD, 2019^[12]). By contrast, policy evaluations focus not only on measuring whether costs are justified in terms of efficiency and value for money (Smismans, 2015^[14]), but also on assessing the extent to which public intervention causes an observed effect, and its relation to intended objectives. Therefore, although these two practices assess public policy programs or activities based on criteria such as efficiency, spending reviews have a specific focus on improving the quality of public expenditure and on proposing reallocations (The World Bank, 2018^[15]). Evaluations can have a broader learning function, helping to assess performance with regard to the policy objectives that were initially fixed. Evaluation can also be a crucial tool to inform the results of spending reviews, without which these may become a purely mechanical exercise.
- **Performance management**, which is defined as a process by which an agency involves its employees, as individuals and members of a group, in improving organisational effectiveness (Walker and Moore, 2011^[16]). This practice, like policy evaluation, aims to increase accountability and provide quality data on a reliable basis to inform decision-making (OECD, 2019^[12]). Nonetheless, although performance management seeks to ensure that a programme is operating as intended in a timely manner and with efficient use of resources, it cannot explain performance variations (Kroll and Moynihan, 2018^[17]). On the contrary, policy evaluation can help make sense of performance outcomes, and create a sense of "policy memory" by taking into account challenges from experiences and good practices that could be incorporated into current performance efforts (Acquah, Lisek and Jacobzone, 2019^[18]; OECD, 2008^[19]).
- **Audit**, aims to 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 will perform their functions effectively, efficiently, ethically and in accordance with the applicable laws and regulations (International Organisation of Supreme Audit Institutions, 2019^[20]). For instance, independent external bodies such as the Supreme Audit Institutions (SAIs) play a role in overseeing and holding government to account for its use of resources, together with the legislature and other oversight bodies. Thus, SAIs support policy evaluation by providing valuable evidence on key government functions, as well as by ensuring accountability (OECD, 2016^[21]), and they can play a double role both as providers of audits, compliance audits and performance audits, as well as of evaluations (see Chapter 3).

How can policy evaluation be defined? Countries' approaches

Governments can benefit from adopting a clear definition of policy evaluation to distinguish it from other practices. Such a definition would also help create a shared understanding within the public sector of the aims, tools and features of policy evaluation. Countries' definitions could therefore include what is policy evaluation, the type of knowledge it should produce, how and why it should be conducted and the actors that are involved.

Figure 1.2 shows that more than half of the survey respondents (27 countries) have adopted a formal definition of policy evaluation: 14 of them have one definition applicable across government, while 13 have several. In terms of OECD countries, a higher share of respondents define policy evaluation (23 of 35), either by having one definition (11) or several (12).

Figure 1.2. Government's formal definition of policy evaluation



Note: n=41 (35 OECD member countries). Kazakhstan answered that they do not know if there is a formal definition for policy evaluation. Answers reflect responses to the question, "Does your government have a formal definition of policy evaluation?" and "Please provide the definition/s and the reference to the relevant documents".
Source: OECD Survey on Policy Evaluation (2018).

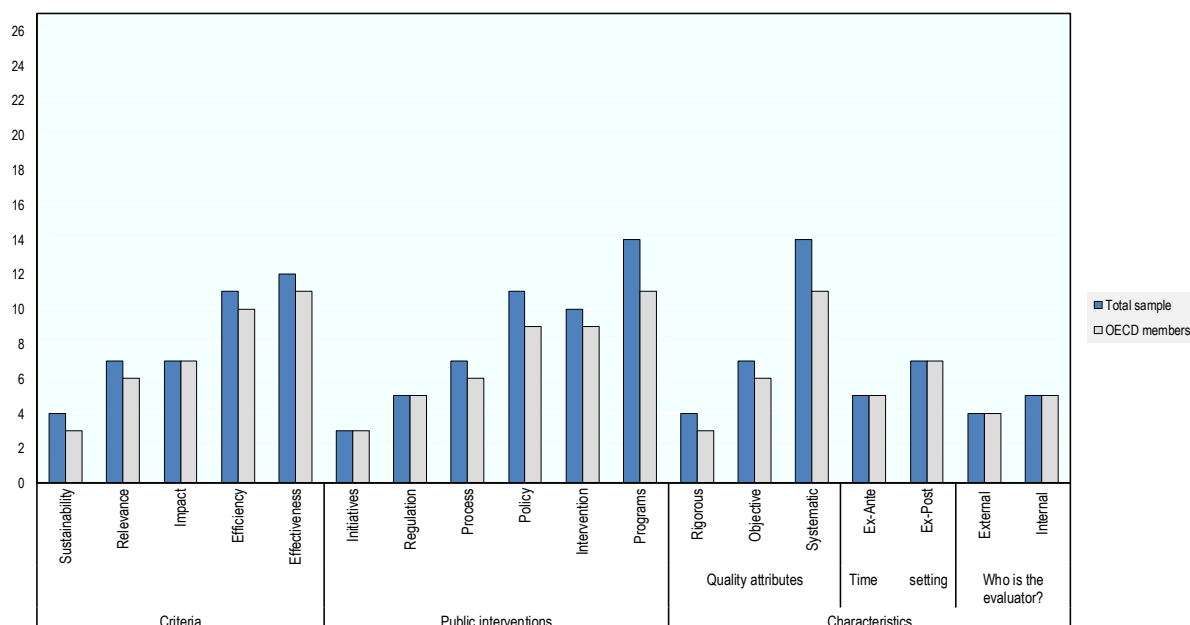
In some cases, the definition is embedded in a legal document. For instance, Japan presents the definition in a law, the Government policy evaluations Act (Act No. 86 of 2001). Argentina defines evaluation in the decree 292/2018, which designates the body responsible for preparing and executing the annual monitoring and evaluation plan for social policies and programmes (2018^[22]). In Latvia, the definition is framed in the development planning system law. Finally, some countries define evaluation in guidelines or manuals as is the case of Mexico (general guidelines for the evaluation of the general public administration programmes (2007^[23])), Costa Rica (manual of evaluation for public interventions (2018^[24])), and Colombia (guide for the evaluation of public policies (2016^[25])).

Key concepts found in country definitions

The analysis for this report has aimed to identify elements of consistency across definitions of evaluation in countries. In fact, while the definitions of policy evaluation diverge across survey respondents, they share several characteristics. The definitions were clustered and mapped across various conceptual dimensions, to highlight elements of a shared understanding.

For the purpose of this report, the key conceptual terms found in the definitions supplied by countries were clustered across three main categories: (1) criteria for evaluation, (2) type of public interventions evaluated, and (3) characteristics (Figure 1.3). This exercise seeks to provide a broad picture of the different approaches for policy evaluation across responding countries, and compare those with practitioners and academic definitions.

Figure 1.3. Conceptual clusters included in the definition of policy evaluation



Note: n=27 (23 OECD member countries). 14 countries (12 OECD member countries) answered they do not have a formal definition for policy evaluation. Moreover, one country (no OECD member country) answered he does not know if there is a formal definition for policy evaluation. Answers reflect affirmative responses to the question, "Does your government have a formal definition of policy evaluation?" and "Please provide the definition/s and the reference to the relevant documents". Aims of policy evaluations.

Source: OECD Survey on Policy Evaluation (2018).

Criteria for policy evaluation

Policy evaluation evaluate different criteria, such as the relevance; effectiveness; efficiency; sustainability and/or impact of a specific intervention (See Box 1.2.).

Box 1.2. Policy evaluation criteria

- **Relevance** — to what extent do the (original) objectives (still) correspond to needs and issues?
- **Effectiveness** — to what extent did a policy/public intervention generate observed effects and changes? To what extent do the observed effects correspond to the objectives?
- **Efficiency** — were the costs involved justified, given the changes and effects achieved?
- **Sustainability** — does the policy/public intervention present net benefits at the long term?
- **Impact** — what are the effects produced by an intervention (i.e. positive or negative, primary and secondary long-term effects produced, directly or indirectly, intended or unintended)?

Source: OECD-DAC (2002^[26]), European Environment Agency (2017^[27]), Smismans, (2015^[14]), and Gasper (2018^[28]).

Most of countries' definitions express that the aim of policy evaluation is to assess the *effectiveness* and *efficiency* of a policy or program. Findings from the last OECD performance budgeting survey² also highlight the interest of countries in evaluating the effectiveness and efficiency of an intervention. Sixteen countries answered that they "usually" evaluate programme effectiveness and efficiency, while seven countries "always" measure effectiveness and five do so for efficiency. This can be explained by governments' needs to identify not which policy options generate the highest impact, but also options that are the most cost-effective (Heider, 2017^[29]).

Seven countries (Austria, Great Britain, Japan, Lithuania, Latvia, Mexico, and Slovakia) refer to impact as criteria. A number of countries directly refer to impact assessment or impact evaluation rather than to evaluation. This could show the misunderstanding of some countries when trying to implement evaluations throughout the policy cycle, but only carrying them out after the implementation of the policy. Seven countries (Australia, Canada, Costa Rica, Great Britain, Japan, México, and Slovakia) also define the aims of policy evaluation in terms of relevance.

Lastly, only Mexico, Slovakia, Great Britain and Costa Rica incorporate the criteria of sustainability in their definitions. One of the reasons for this can be the lack of a stringent and clear definition of sustainability and its association with environmental measurements. According to OECD-DAC evaluation criteria (2002^[26]), sustainability includes the “examination of the financial, economic, social, environmental, and institutional capacities of the systems needed to sustain net benefits over time”. Another explanation may be that these results indicate that sustainability may be less of a pressing challenge for sampled countries (mostly OECD countries).

Types of public interventions set out in the definitions

Regarding which kind of public interventions are present in the different definitions, countries generally focus on programmes, interventions and policies. Some of them additionally consider activities such as regulations and processes, adopting a wider definition of policy evaluation. These findings may also demonstrate some conflation between monitoring and evaluation terms, such as referring explicitly to ongoing operational decision-making (processes). In general, survey answers reveal a countries’ relative difficulty in defining the concept ‘policy’, as respondents refer to interventions, programmes and initiatives as falling under that category. A good example of definition, where they are clearly differentiate between different types of public interventions, is the definition of Costa Rica (see Box 1.3).

Box 1.3. Public interventions: Costa Rica

In its manual for the evaluation of public interventions (2018^[24]), the *Ministry of National Planning and Economic Policy* from Costa Rica differentiates **four different types of public interventions** (policy, plan, program and project), depending on the characteristics of the problem to be addressed (magnitude of the problem, resources available to respond to it, scope, target population, etc.). This classification can be summarised as follows:

- Policies: Defined course of action to guide or achieve an objective, expressed in guidelines, strategic aims and actions on a specific topic.
- Plans: Integrated set of programmes that respond to the fulfilment of objectives and goals, which are executed in the short, medium and long term, and include dedicated resources.
- Programmes: Set of interrelated projects that look to achieve specific and common objectives.
- Projects: A set of activities that aim to achieve specific objectives, with a given budget and by a specific date, mainly oriented to the production of goods and services.

Source: Ministry of National Planning and Economic Policy of Costa Rica (2018^[24]).

Key characteristics of policy evaluation should be the evaluator

Definitions can also include **quality attributes** (systematic, objective, and rigorous); **time setting**, “when” (Ex-post vs. Ex-ante) and **who** (internal evaluation vs. external evaluation).

As will be explained in the chapter on *Quality and use of policy evaluations*, the **quality of evaluations** depends on both their methodological rigor and their trustworthiness. Reflecting both these aspects, the most common characteristics found in countries’ definitions relating to quality are the following:

- **Systematic:** An evaluation should be carried out using a planned and organised procedure or an agreed set of methods. For instance, decisions need to be based on systematic approaches (i.e. theory-based approach: logic, reasoning, and by an accurate guide or principle) instead of on unfounded assumptions (Gasper, 2018^[30]).
- **Rigour:** evaluations should be developed using well-designed and well-implemented methods tailored to the target question (Commission on Evidence-Based Policymaking, 2017^[31]).
- **Objective:** An evaluation should be conducted from an impartial position, without any personal or political factor influencing (from a researcher or policymaker) the research design and its implementation (Parkhurst, 2017^[32]).

According to the survey (Figure 1.3), 15 of the respondents (12 OECD countries) mentioned at least one of the terms above in their definition. A majority of them include the fact that policy evaluation must be *systematic* (14 of all the survey respondents and 11 OECD countries). Only seven countries of the total respondents (from which six are OECD countries) specify a policy evaluation must be *objective*. Canada and Spain take into account characteristics similar to objectiveness such as *neutral* (in Canada definition), and *reasoned* (in Spain definition). Only Great Britain, Japan, Argentina and Mexico mention *rigorous* (see further examples in Box 1.4).

Box 1.4. Quality attributes in countries definitions

Argentina: “The evaluation of policies, programmes, plans and projects with social impact, comes from a form of applied, *systematic*, planned and *rigorous* social research; aimed at identifying, obtaining and providing data and valid and reliable information about them; which will allow improving both its design and its implementation, and ensure access to the human rights it seeks to promote (Decree 292/2018)”.

Mexico: “Evaluation is a *systematic* and *objective* analysis of federal programs whose purpose is to determine the relevance and achievement of its objectives and goals, as well as its efficiency, quality, results, impact and sustainability”.

Lithuania: “Evaluation is a *systematic* and *objective* determination of the suitability, effectiveness, efficiency, usefulness and long-term impact of the planned, executed or completed programs” (Resolution on Strategic Planning methodology No 827 approved on 2002 June 6).

Source: Poder Ejecutivo Nacional de Argentina (2018^[33]) and Consejo Nacional de Evaluación de la Política de Desarrollo Social de México (2007^[23]).

The **time setting** (ex-ante vs. ex-post) is a key criterion for differentiation, both from an analytical and methodological perspective. The term *ex-post evaluation* refers to a retrospective evaluation that can be interim (i.e. at the mid-term of an initiative), final (at its conclusion), or *ex post* in the strict sense (placed several years after the intervention has finished) (Smismans, 2015^[34]). *Ex post* evaluation can be a tool for accountability, also close to performance audit and control, for example when it is performed through supreme audit institutions, or internal inspection bodies. Yet, *ex post* evaluation is also important to facilitate learning, to understand if and when the objectives of policies were attained and spell out a theory of change. *Ex-ante evaluation* on the other hand refers to a set of rules, instructions and procedures that enable public institutions to have a portfolio of socially profitable

investment initiatives before their implementation (Ministerio de Desarrollo Social y Familia de Chile, 2019^[35]). It provides an assessment whether the strategy and objectives proposed are relevant to target population needs; and whether the assumptions concerning expected results and impacts are realistic and in line with the resources available (The European Network for Rural Development, 2014^[36]).

New approaches in policy evaluation are trying to disentangle the challenges of linking *ex-ante* and *ex-post* appraisal, and apply a policy evaluation system focused on the entire policy cycle (thus covering these two types of evaluations), in all the policy areas (Mergaert and Minto, 2015^[37]; Smismans, 2015^[34]). For survey respondents, the policy evaluation definition is more commonly linked to the second term (*ex-post*), related to an already implemented policy. Seven OECD countries of the survey respondents refer to *ex-post*, and only five OECD countries as well to *ex-ante*. Specific countries defy the norm such as Norway which specifies that evaluations can be undertaken prior to (*ex-ante*), during, or after implementation (*ex-post*) (2006^[38]). This also depends on the context and use of policy evaluation. Thus, while evaluation of public expenditure, policies and programmes tends to be overwhelmingly *ex-post*, by contrast, in the regulatory area, the focus is most often on *ex-ante* evaluation of regulations.

The third common characteristic is related to **“who” carries out an evaluation**: *external evaluations* (also known as “informal”, “outside” or “society-driven” evaluation) or *internal evaluations* (also known as “formal”, “inside” or “government-driven” evaluation) (Schoenefeld and Jordan, 2017^[39]; Weiss, 1993^[40]; Hildén, 2014^[41]). Survey findings (See Figure 1.3) report that only four OECD countries of the total respondents mention external evaluations in their definitions, and five OECD countries internal evaluation.

External evaluation refers to an evaluation of an intervention conducted by entities and/or individuals outside the government. This type of evaluation could be considered to be more independent, it can take a more critical look at the policy being studied and their results can be potentially more trusted (Schoenefeld and Jordan, 2017^[39]). However, as much as this type of evaluation could be independent from government actions, it can also be limited by the knowledge of the evaluator about the context and political process, as well as access to relevant data.

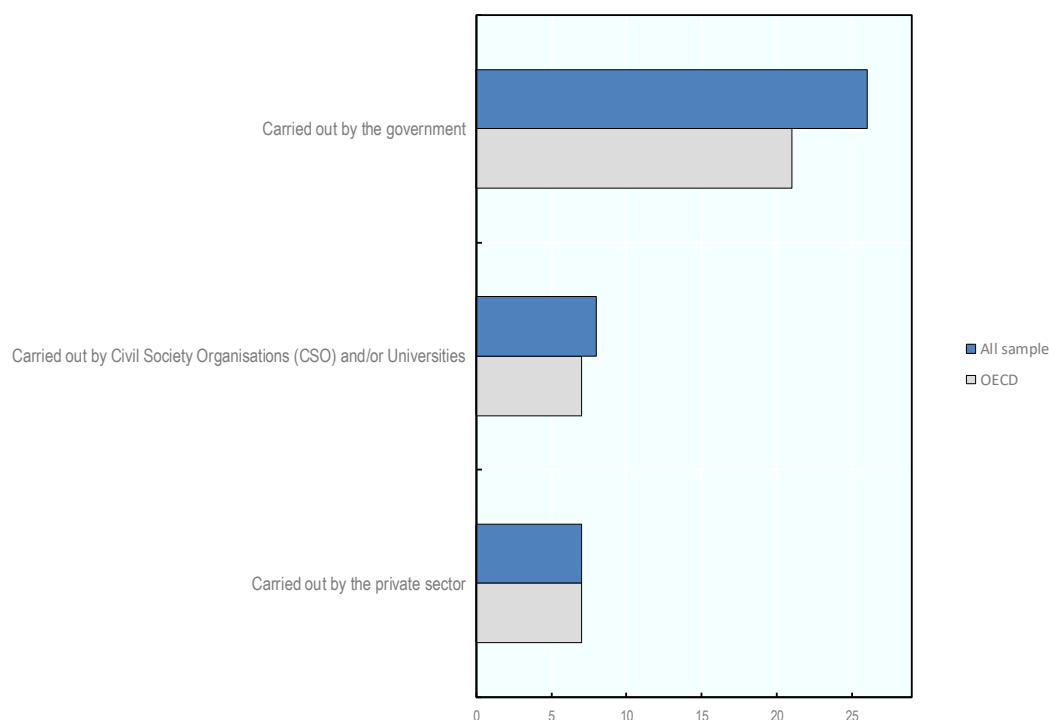
The notion of *internal evaluation* refers to an evaluation of a development intervention conducted by a governmental institution. Internal evaluators may have more knowledge about a public policies, provide a more accurate assessment according to local contexts and have easier access to inside data than what an external evaluator could supply (Weiss, 1993^[40]; Schoenefeld and Jordan, 2017^[39]). Nonetheless, in the process of conducting an evaluation, the internal evaluators can be under political pressure and time constraints to show good results, which can affect the validity of the findings of the evaluation and its public deliberation.

Overall, there is still no evidence to determine which type of evaluation is better or preferred. In practice, the differences between internal and external evaluations can sometimes be blurred as hybrid approaches, mixing internal evaluations combined with some external evaluation for specific or more technical aspects of a policy or programme. Moreover, a government can commission the evaluation to an external organisation (e.g. NGO, universities), while still ensuring that civil servants control the research questions addressed by the evaluation (i.e. principle agent relationships (Schoenefeld and Jordan, 2017^[39]).

The selection of “internal vs. external” evaluations will depend on how each approach fits with the goals of the policy evaluation and the overall social-political circumstances (Schoenefeld and Jordan, 2017^[39]). Some countries have adopted clear criteria for determined which approach fits best in what circumstances. For example, the Cabinet Implementation Unit from Australia (2014^[42]) specifies that the availability of resources and capacity will determine whether the evaluation is conducted internally or externally.

The concept of external evaluation also covers a variety of actors. Figure 1.4 presents an overview of what type of actors typically carry out evaluations in surveyed countries. For instance, although the evaluation of government-wide policy priorities (GWPP) are still mainly carried out by the government (26 countries of all the respondents, and 21 OECD countries), eight countries choose to commission evaluations to civil society organisations or universities and seven countries choose the private sector.

Figure 1.4. Actors involved on carrying out evaluations of government-wide policy priorities (GWPP)



Note: n=29 (24 OECD member countries). Four countries (all OECD member countries) answered that they do not have government-wide policy priorities. Moreover, nine countries (7 OECD member countries) answered that they do not evaluate their government-wide policy priorities. Answers reflect affirmative responses to the question, "Evaluations of government-wide policy priorities are carried out by ". The option "Other" is not included.

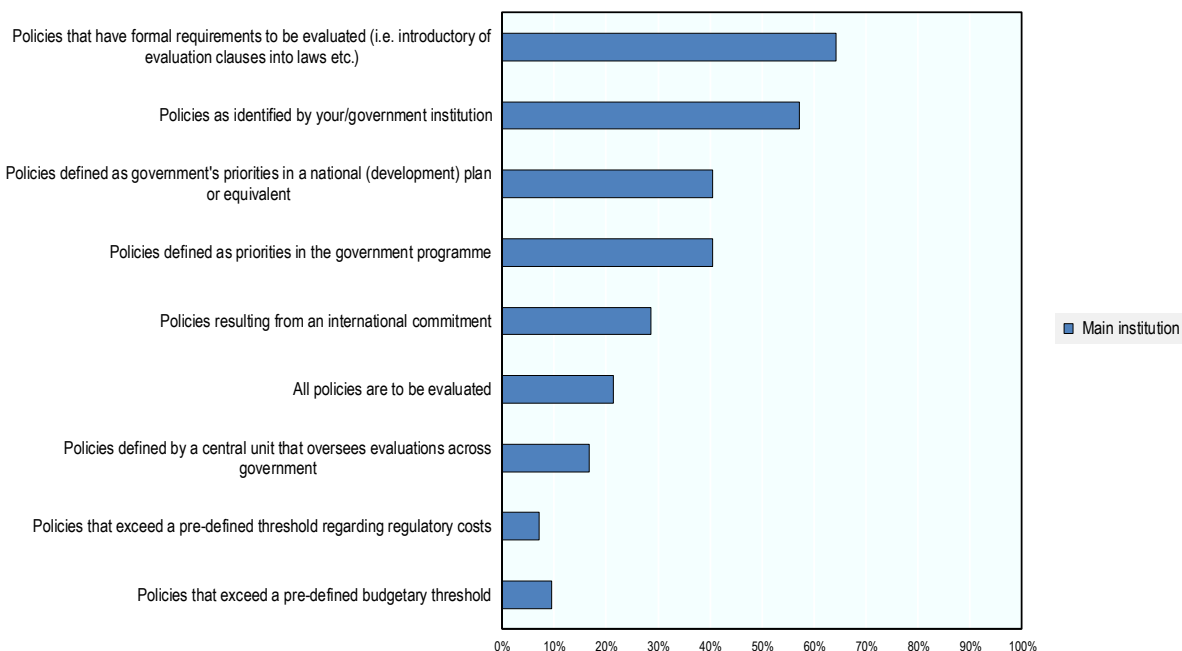
Source: OECD Survey on Policy Evaluation (2018).

At the sector level, ministries stand out as carrying out in average about 53% of their evaluations internally: either by the own institution (26 countries in Health and 17 in the Public Sector Reform (PSR)) or by a central unit (15 countries in Health and 5 in PSR). Ministries only do external evaluations around 36% of cases: either with civil society organisations or universities (12 countries in Health and 8 in PSR) or with the private sector (13 countries in Health and 10 in PSR).

What policies are being evaluated?

According to Figure 1.5, most of the countries evaluate policies that have formal requirements, such as policies with evaluation clauses into laws, policies identified by government institutions and policies defined as government priorities in a national plan or program. The influence of international commitments can be seen in some countries, such as Austria, Germany, Finland, Spain, and Greece. Lastly, few countries evaluate all policies; which can be due to an evaluation can be time consuming and entails a costly process.

Figure 1.5. What policies are being evaluated?



Note: n=42 (35 OECD member countries). Answers reflect responses to the question, "Which policies are evaluated?". The option "Other" is not included.

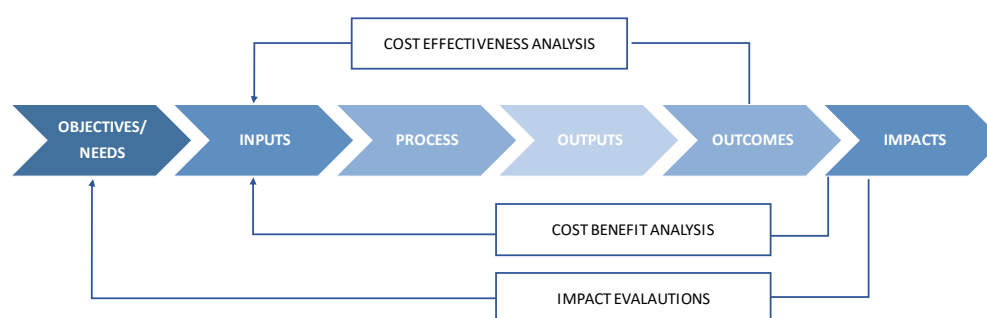
Source: OECD Survey on Policy Evaluation (2018).

Methodologies and tools used in policy evaluation

All phases of the policy chain can be evaluated through different types of policy evaluation (see Figure 1.6). This report defines the different stages of the policy chain as follows:

1. The input level refers to the resources employed to implement a policy (OECD, 2016^[43]) such as staff, money, time, equipment, etc.
2. The process level refers to the activities that were undertaken in a policy (OECD, 2016^[43]).
3. The output level refers to a first level of results, directly associated with the products delivered by the policy implemented.
4. The outcome level refers to the medium-term (directly) consequences of the policy implemented (OECD, 2016^[43]).
5. The impact level looks at the long-term consequence of a policy initiative (OECD, 2016^[43]).

Figure 1.6. Policy evaluation criteria along the policy chain



Source: OECD

The chain of “input-activities-output-outcome-impact” and its causality is particularly complex. Evaluating a single initiative, versus a comprehensive action plan, requires different tools and can probably reach different levels of understanding (OECD, 2016^[43]). As shown in Figure 1.6, each of these elements can be compared to one another to evaluate different aspects of a policy/public intervention. For instance, a cost effectiveness analysis will require the comparison between the cost of an intervention (from the inputs or resources employed) and the outcomes obtained. Table 1.2 presents in more detail some of the evaluations and methodologies used to respond these target questions.

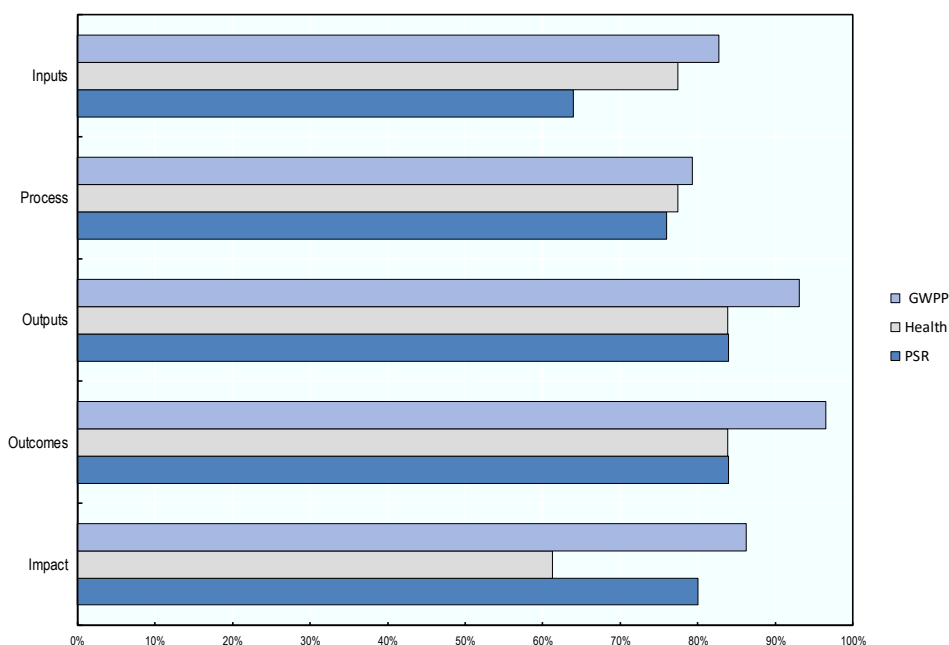
Table 1.2. Type of evaluations

Stages	Target question	Type of evaluations
Process	How was the policy delivery?	Process evaluations is “the analysis of what has sometimes been called the “intervention logic” or causal chain of programmes” (Vammalle and Ruiz Rivadeneira, 2017 ^[44]). “Questions might, for instance, seek to describe how individuals were recruited onto the scheme, what criteria were used to recruit them, and what the qualifications of training providers were. It might explore to what extent these factors varied across different parts of the country, and whether recruitment processes operated in favour of or to the detriment of particular groups” (HM Treasury, 2011 ^[9])
Outcomes	Has the target population of the program received the services/product? Was the policy justified? Did the benefits overcome the costs? (efficiency)	Outcome Evaluation measures program or policy effects in the target population by assessing the progress towards achieving the outcomes that the program or policy is aiming to deliver Economic evaluations show whether those outcomes justified that policy, including whether the costs of the policy have been outweighed by the benefits (HM Treasury, 2011 ^[9]) Cost analysis: is used to determine the cost of implementing a policy or program (Crowley et al., 2018 ^[45]). Cost effectiveness analysis: Focus on the cost of the inputs and outcomes achieved in the intervention. It is also known as a way of comparing the costs of two or more interventions to reduce or produce a single beneficial outcome (Crowley et al., 2018 ^[45]). Cost Benefits Analysis: consists in a method in which both costs and outcomes of an intervention are valued in monetary terms, permitting a direct comparison of the benefits produced by the intervention (same metrics e.g. dollars) (Steuerle and Jackson, 2016 ^[46] ; OECD, 2018 ^[47]).
Impacts	Does a policy work? (effectiveness)	Impact evaluations seek to answer to the question “Does a policy work?” Those effects could be positive or negative, primary or secondary intended or unintended, direct or indirect (OECD, 2010 ^[48]). This type of study seeks to determine the efficacy and effectiveness of a policy or program, with a counterfactual control group to understand what would happen to a population if a specific policy or programme were not implemented (Morton, 2009 ^[49]).

Source: Crowley et al. (2018^[45]), Flay et al (2005^[50]), Morton (2009^[49]), HM Treasury (2011^[9]), OECD (2010^[48]) (2018^[47]), and Steuerle and Jackson (2016^[46]).

Survey data (see Figure 1.7) shows that on average the most evaluated elements in government-wide policy priorities (GWPP) and in policies on charge of ministries of Health and PSR are outputs (87%) and outcomes (88%), followed by process (78%), impact (76%), and inputs (75%). On the other hand, there is a major variance in the *impact* element. Impacts are more commonly evaluated for GWPP (86% in all the survey respondents and 88% in OECD countries) and for the policies of PSR ministries (80% in all the survey respondents and 85% in OECD countries), compared to the policies of Health ministries (61% in all and OECD countries respondents).

Figure 1.7. Elements in the policy cycle chain that are evaluated



Note: The chart is expressed as a percentage of responding countries as number of respondents differ for the main institution, health and PSR. n=29 (24 OECD member countries). 4 countries (all OECD member countries) answered that they do not have government-wide policy priorities. Moreover, 9 countries (7 OECD member countries) answered that they do not evaluate their government-wide policy priorities. For the Health ministries n=31 (28 OECD member countries). 9 countries (7 OECD member countries) did not participate on this survey. Moreover, 2 countries (1 OECD member country) are not included as they answered that none of the policies that fall in their institution's responsibility are evaluated. For the PSR ministries n=25 (20 OECD member countries). 11 countries (10 OECD member countries) did not participate on this survey. Moreover, 6 countries (5 OECD member countries) are not included as they answered that none of the policies that fall in their institution's responsibility are evaluated. Answers reflect responses to the question, "Which elements are evaluated by your institution? (Check all that apply)".

Source: OECD Survey on Policy Evaluation (2018).

More generally, given that evaluations are a costly activity, it is important to justify the need for evaluation and the resources that it will require, which would call governments to establish some sets of criteria for determining when and what type of evaluation is needed. Setting threshold and proportionality criteria is something already well embedded in some countries concerning the Regulatory Impact Assessment process.

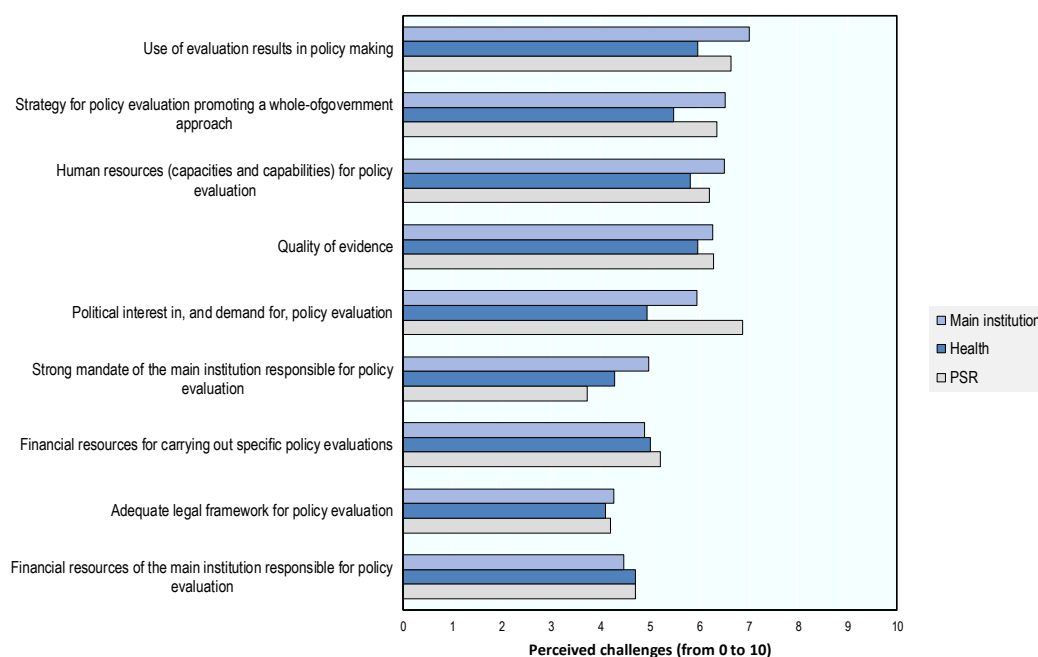
This may also explain why, despite the potential of randomised control trials (RCT) or quasi-experimental designs (QED) to provide rigorous findings, the evidence shows that they are comparatively less used (see further details in). In addition, these results could be related to the not always practical use of RCTs due to legal, political and ethical considerations such as ensuring participant selection procedures are fair, there is an acceptable balance of benefits and harms, and participants provide an informed consent (Acquah, Lisek and Jacobzone, 2019^[18]; Goldstein et al., 2018^[51]).

Towards sound policy evaluation systems: promoting institutionalisation, quality and use

Despite the growing interest for and acknowledgment of its contribution to improve the design and implementation of public policies, policy evaluation often constitutes the weakest link in the policy cycle. The reasons for this are manifold. Firstly, some countries face *technical barriers for carrying out evaluations* such as the challenges of governments to create and share verifiable, accurate, useable and unbiased data within and outside public administration (Rutter, 2012^[52]).

Findings from the OECD survey suggest (see Figure 1.8) that the four main challenges for promoting policy evaluation across government are: the limited use of evaluation results in policy-making, the absence of a strategy for policy evaluation that promotes a whole-of-government approach, the limited availability of human resources (capacities and capabilities) for policy evaluation. Ministries perceive issues related to the quality of the evidence, and related to the political interest in, demand for, policy evaluation.

Figure 1.8. Challenges for promoting policy evaluation across government



Note: For the main institution n=42 (35 OECD member countries). For the Health ministries n=31 (28 OECD member countries). 9 countries (7 OECD member countries) did not participate on this survey. Moreover, 2 countries (1 OECD member country) are not included as they answered that none of the policies that fall in their institution's responsibility are evaluated. For the PSR ministries n=25 (20 OECD member countries). 11 countries (10 OECD member countries) did not participate on this survey. Moreover, 6 countries (5 OECD member countries) are not included as they answered that none of the policies that fall in their institution's responsibility are evaluated. Answers reflect responses to the questions, "What are the government's current challenges for promoting policy evaluations?" for the main institution and "What are current challenges for promoting policy evaluation in your institution?" for Health and PSR, where 0 indicates that is a "rare challenge", 5 is "Neutral", and 10 is a "principal challenge".

Source: OECD Survey on Policy Evaluation (2018).

These four main challenges perceived by countries can be considered – to a certain extent and depending of the institutional context – as inter-dependents. The lack of human resources in terms of capabilities and capacities (for instance to commission or undertake evaluations) can probably affect the quality of the evaluations in a negative way. As will be analysed in the subsequent chapters, the quality of evaluations might also influence the use of its results. Moreover, considering these elements, the development an integrated and whole-of-government strategy to promote policy evaluation is not an easy task, due to the aforementioned limited capacities and capabilities, the lack of political support, and probably also to the absence of analytical frameworks to develop such strategies. Because of these interdependencies, this report adopts a systematic approach of the promotion of policy evaluation within governments.

Beyond these challenges, another key element is the issue of the timeliness of the evaluation results, that is, whether they arrive at the time of decision-making. Usually, the time span of Ministers and political life implies a very short lead-time to make decisions. Evaluations, on the other hand, require time. This challenge may call for two kinds of reactions. The first is a rapid adaptive response, doing a quick evaluation within the available timeframe, and working out proxies and similar studies. Another approach is to invest upfront, and to have a certain reserve of “evaluation capacity” and evaluative studies that can be ready for when the demand arises. These allow to draw on the existing stock of knowledge and to provide answers to the short-term demands when it arrives. However, this requires an investment *ex ante*, the capacity to manage the stock of knowledge in a strategic manner in light of anticipated demand. This issue is crucial for establishing a well-functioning evaluation system.

The value of a systemic whole-of-government approach

This report promotes a systems’ approach to policy evaluation. A system can be defined as “elements linked together by dynamics that produce an effect, create a whole new system or influence its elements” (OECD, 2017^[53]). A policy evaluation system can be defined, following Lazaro, as: one in which evaluation is a regular part of the life cycle of public policies and programmes, it is conducted in a methodologically rigorous and systematic manner, in which its results are used by political decision-makers and managers, and those results are also made available to the public” (2015^[54]). Thus, a system calls for constant adjustment throughout the policy cycle, with implications for the ways in which institutions, processes, skills and actors are organised (OECD, 2017^[53]).

In order to develop and/or implement a strategy for promoting a whole-of-government approach on policy evaluation, this report adopts a tiered approach toward through the triple lens of **institutionalisation, quality and use** (see Box 1.5)

Box 1.5. Components of a sound policy evaluation system

This report approaches the idea of a sound policy evaluation system through the following three dimensions:

- Institutionalisation: the systematic process of embedding evaluation practices into more formal and systematic approaches. It can include establishing an evaluation system in governmental settings through specific policies or strategies (Lázaro, 2015^[54]; Gaarder and Briceño, 2010^[55]).
- Quality: defined as policy evaluations that are technically rigorous as well as well governed; that is be independent and appropriate for the decision-making process (Picciotto, 2013^[56])
- Use: which is defined under three conditions (Ledermann, 2012^[57]):
 - *Symbolic use* (also known as persuasive), occurs when the results of evaluations are taken up to justify or legitimise a pre-existing position, without changing it;
 - *Conceptual use* happens when evaluation results lead to an improved understanding or a change in the conception of the subject of evaluation;
 - *Instrumental use* is when evaluation recommendations inform decision-making and lead to an alteration in the object of evaluation.

Source: Lázaro (2015^[54]), Gaarder and Briceño (2010^[55]), Ledermann (2012^[57]), and Picciotto (2013^[56]).

Note

¹ This definition is adapted from the Open Government: The Global Context and the Way Forward (2016^[43]), which is based on “OECD-DAC Glossary” in Guidelines for Project and Programme Evaluation (2009^[84])

References

- 115th Congress (2019), *Public Law No: 115-435 (01/14/2019) - Foundations for Evidence-Based Policymaking Act of 2018*, <https://www.congress.gov/bill/115th-congress/house-bill/4174>. [7]
- Acquah, D., K. Lisek and S. Jacobzone (2019), “The Role of Evidence Informed Policy Making in Delivering on Performance: Social Investment in New Zealand”, *OECD Journal on Budgeting*, Vol. 19/1, <https://dx.doi.org/10.1787/74fa8447-en>. [18]
- Canada Treasury Board (2016), *Policy on results*, <https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=31300> (accessed on 20 September 2019). [8]

- Commission on Evidence-Based Policymaking (2017), *The Promise of Evidence-Based Policymaking: Report of the Commission on Evidence-Based Policymaking*, <https://www.cep.gov/report/cep-final-report.pdf> (accessed on 6 August 2019). [31]
- CONEVAL (2007), *Lineamientos generales para la evaluación de los Programas Federales de la Administración Pública Federal*, https://www.coneval.org.mx/rw/resource/coneval/eval_mon/361.pdf (accessed on 18 June 2019). [23]
- Crowley, D. et al. (2018), “Standards of Evidence for Conducting and Reporting Economic Evaluations in Prevention Science”, *Prevention Science*, Vol. 19/3, pp. 366-390, <http://dx.doi.org/10.1007/s11121-017-0858-1>. [45]
- Departamento Nacional de Planeación (2016), *¿Qué es una Evaluación?*, <https://sinergia.dnp.gov.co/Paginas/Internas/Evaluaciones/%C2%BFQu%C3%A9-es-Evaluaciones.aspx>. [25]
- Department of the Prime Minister and Cabinet (2014), *Guide to Implementation Planning*, Australian government, <https://www.pmc.gov.au/sites/default/files/files/guide-to-implementation-planning.pdf> (accessed on 12 July 2019). [42]
- European Environment Agency (2017), “EEA guidance document-policy evaluation”, <https://www.researchgate.net/publication/317594615>. [27]
- Flay, B. et al. (2005), “Standards of Evidence: Criteria for Efficacy, Effectiveness and Dissemination”, *Prevention Science*, Vol. 6/3, pp. 151-175, <http://dx.doi.org/10.1007/s11121-005-5553-y>. [50]
- Gaarder, M. and B. Briceño (2010), “Institutionalisation of government evaluation: balancing trade-offs”, *Journal of Development Effectiveness*, Vol. 2/3, pp. 289-309, <http://dx.doi.org/10.1080/19439342.2010.505027>. [55]
- Gasper, D. (2018), “Policy Evaluation: From Managerialism and Econocracy to a Governance Perspective”, in *International Development Governance*, Routledge, <http://dx.doi.org/10.4324/9781315092577-37>. [28]
- Gasper, D. (2018), “Policy Evaluation: From Managerialism and Econocracy to a Governance Perspective”, in *International Development Governance*, Routledge, <http://dx.doi.org/10.4324/9781315092577-37>. [30]
- Goldstein, C. et al. (2018), *Ethical issues in pragmatic randomized controlled trials: A review of the recent literature identifies gaps in ethical argumentation*, BioMed Central Ltd., <http://dx.doi.org/10.1186/s12910-018-0253-x>. [51]
- Heider, C. (2017), *Rethinking Evaluation - Efficiency, Efficiency, Efficiency*, <https://ieg.worldbankgroup.org/blog/rethinking-evaluation-efficiency>. [29]
- Hildén, M. (2014), “Evaluation, assessment, and policy innovation: exploring the links in relation to emissions trading”, *Environmental Politics*, Vol. 23/5, pp. 839-859, <http://dx.doi.org/10.1080/09644016.2014.924199>. [41]
- HM Treasury (2011), *The Magenta Book: Guidance for evaluation*, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/220542/magenta_book_combined.pdf (accessed on 18 June 2019). [9]

- Howlett, M. (2019), "Policy analytical capacity and evidence-based policy-making: Lessons from Canada", https://doi.org/10.1111/j.1754-7121.2009.00070_1.x. [3]
- International Organisation of Supreme Audit Institutions (2019), *ISSAI 100: Fundamental Principles of Public-Sector Auditing*, <http://www.issai.orghttp://www.intosai.org> (accessed on 8 January 2020). [20]
- Kroll, A. and D. Moynihan (2018), "The Design and Practice of Integrating Evidence: Connecting Performance Management with Program Evaluation", *Public Administration Review*, Vol. 78/2, pp. 183-194, <http://dx.doi.org/10.1111/puar.12865>. [17]
- Lazaro, B. (2015), *Comparative Study on the Institutionalization of Evaluation in Europe and Latin America*, Eurosocal Programme. [5]
- Lázaro, B. (2015), *Comparative study on the institutionalisation of evaluation in Europe and Latin America*, Eurosocal Programme, Madrid, http://sia.eurosocal-ii.eu/files/docs/1456851768-E_15_ENfin.pdf (accessed on 9 July 2019). [54]
- Ledermann, S. (2012), "Exploring the Necessary Conditions for Evaluation Use in Program Change", *American Journal of Evaluation*, Vol. 33/2, pp. 159-178, <http://dx.doi.org/10.1177/1098214011411573>. [57]
- McDavid, J., I. Huse and L. Hawthorn (2006), *Program Evaluation and Performance Measurement: An Introduction to Practice*, <https://study.sagepub.com/mcdavid3e> (accessed on 28 January 2020). [10]
- Mergaert, L. and R. Minto (2015), "Ex Ante and Ex Post Evaluations: Two Sides of the Same Coin? The Case of Gender Mainstreaming in EU Research Policy", *Symposium on Policy Evaluation in the EU*, <http://dx.doi.org/10.1017/S1867299X0000427X>. [37]
- Ministerio de Desarrollo Social y Familia de Chile (2019), *Evaluación Social Ex Ante*, <http://sni.ministeriodesarrollosocial.gob.cl/evaluacion-iniciativas-de-inversion/evaluacion-ex-ante/> (accessed on 5 July 2019). [35]
- Ministerio de Planificación Nacional y Política Económica (2018), *Manual de Evaluación para Intervenciones Públicas*, <https://documentos.mideplan.go.cr/share/s/6eepeLCEsrKkft6Mf5SToA> (accessed on 5 August 2019). [24]
- Ministry of Finance (2006), *The Norwegian Government Agency for Financial Management (DFØ)*, <https://www.regjeringen.no/en/dep/fin/about-the-ministry/etater-og-virksomheter-under-finansdepartementet/subordinateagencies/the-norwegian-government-agency-for-fina/id270409/> (accessed on 12 July 2019). [38]
- Ministry of Finance of The Netherlands (2018), *Arrangements for periodic evaluation research*, <https://wetten.overheid.nl/BWBR0040754/2018-03-27> (accessed on 12 July 2019). [6]
- Morton, M. (2009), *Applicability of Impact Evaluation to Cohesion Policy 1 Report Working Paper of*, https://ec.europa.eu/regional_policy/archive/policy/future/pdf/4_morton_final-formatted.pdf (accessed on 8 August 2019). [49]
- OECD (2019), *Budgeting and Public Expenditures in OECD Countries 2019*, OECD Publishing, Paris, <https://dx.doi.org/10.1787/9789264307957-en>. [12]

- OECD (2019), *Open Government in Biscay*, OECD Public Governance Reviews, OECD Publishing, Paris, <https://dx.doi.org/10.1787/e4e1a40c-en>. [11]
- OECD (2018), *Building Capacity for Evidence Informed Policy Making: Towards a Baseline Skill Set*, <http://www.oecd.org/gov/building-capacity-for-evidence-informed-policymaking.pdf> (accessed on 3 September 2019). [1]
- OECD (2018), *Cost-Benefit Analysis and the Environment: Further Developments and Policy Use*, OECD Publishing, Paris, <https://dx.doi.org/10.1787/9789264085169-en>. [47]
- OECD (2018), *Draft Policy Framework on Sound Public Governance*, <http://www.oecd.org/gov/draft-policy-framework-on-sound-public-governance.pdf> (accessed on 8 July 2019). [2]
- OECD (2018), *OECD Performance Budgeting Survey*. [59]
- OECD (2017), *Government at a Glance - OECD*, OECD, Paris, <http://www.oecd.org/gov/govataglace.htm> (accessed on 9 July 2019). [4]
- OECD (2017), *Systems Approaches to Public Sector Challenges: Working with Change*, OECD Publishing, Paris, <https://dx.doi.org/10.1787/9789264279865-en>. [53]
- OECD (2016), *Open Government: The Global Context and the Way Forward*, OECD Publishing, Paris, <https://dx.doi.org/10.1787/9789264268104-en>. [43]
- OECD (2016), *Supreme Audit Institutions and Good Governance: Oversight, Insight and Foresight*, OECD Public Governance Reviews, OECD Publishing, Paris, <https://doi.org/10.1787/9789264263871-en>. [21]
- OECD (2011), "Typology and implementation of spending reviews", *OECD SBO Meeting on Performance and Results*, [http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=GOV/PGC/SBO\(2011\)9&doclanguage=en](http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=GOV/PGC/SBO(2011)9&doclanguage=en) (accessed on 1 August 2019). [13]
- OECD (2010), "DAC Guidelines and Reference Series: Quality Standards for Development Evaluation", <https://www.oecd.org/development/evaluation/qualitystandards.pdf> (accessed on 9 July 2019). [48]
- OECD (2008), *Performance Budgeting: A Users' Guide*, <https://www.oecd.org/gov/budgeting/Performance-Budgeting-Guide.pdf> (accessed on 2 August 2019). [19]
- OECD-DAC (2009), "Guidelines for Project and Programme Evaluations", https://www.entwicklung.at/fileadmin/user_upload/Dokumente/Projektentwicklung/Englisch/Guidelines_for_Project_and_Programme_Evaluations.PDF (accessed on 20 September 2019). [58]
- OECD-DAC (2002), *Glossary of Key Terms in Evaluation and Results Based Management*, <https://www.oecd.org/dac/evaluation/18074294.pdf> (accessed on 18 June 2019). [26]
- Parkhurst, J. (2017), *The politics of evidence : from evidence-based policy to the good governance of evidence*, Routledge, London, <http://researchonline.lshtm.ac.uk/3298900/> (accessed on 23 November 2018). [32]
- Picciotto, R. (2013), "Evaluation Independence in Organizations", *Journal of MultiDisciplinary* [56]

- Evaluation*, Vol. 9/20, p. 15.
- Poder Ejecutivo Nacional (2018), *Decreto 292/2018: Evaluación de Políticas y Programas Sociales*, 11-04-2018, <http://servicios.infoleg.gob.ar/infolegInternet/verNorma.do?id=308653> (accessed on 8 July 2019). [33]
- Poder Ejecutivo Nacional de Argentina (2018), *Decreto 292/2018: Evaluación de Políticas y Programas Sociales*. [22]
- Rutter, J. (2012), *Evidence and Evaluation in Policy making*, Institute for Government, https://www.instituteforgovernment.org.uk/sites/default/files/publications/evidence%20and%20evaluation%20in%20template_final_0.pdf. [52]
- Schoenefeld, J. and A. Jordan (2017), "Governing policy evaluation? Towards a new typology", *Evaluation*, Vol. 23/3, pp. 274-293, <http://dx.doi.org/10.1177/1356389017715366>. [39]
- Smismans, S. (2015), "Policy Evaluation in the EU: The Challenges of Linking Ex Ante and Ex Post Appraisal", *Symposium on Policy Evaluation in the EU*, <http://dx.doi.org/10.1017/S1867299X00004244>. [14]
- Smismans, S. (2015), "Policy Evaluation in the EU: The Challenges of Linking Ex Ante and Ex Post Appraisal", *Symposium on Policy Evaluation in the EU*, <http://dx.doi.org/10.1017/S1867299X00004244>. [34]
- Steuerle, E. and L. Jackson (2016), *Advancing the power of economic evidence to inform investments in children, youth, and families*, National Academies Press, <http://dx.doi.org/10.17226/23481>. [46]
- The European Network for Rural Development (2014), *The Ex Ante Evaluation 2014-2020 RDPs*, https://enrd.ec.europa.eu/evaluation/publications/guidelines-ex-ante-evaluation-2014-2020-rdps_en (accessed on 29 January 2020). [36]
- The World Bank (2018), *Spending Review Manual: Bulgaria*. [15]
- Vammalle, C. and A. Ruiz Rivadeneira (2017), "Budgeting in Chile", *OECD Journal on Budgeting*, Vol. 16/3, <https://dx.doi.org/10.1787/budget-16-5jfw22b3c0r3>. [44]
- Walker, K. and K. Moore (2011), *Performance Management and Evaluation: What's the difference?*, <https://www.childtrends.org/wp-content/uploads/2013/06/2011-02PerformMgmt.pdf> (accessed on 18 July 2019). [16]
- Weiss, C. (1993), "Where politics and evaluation research meet", *Evaluation Practice*, Vol. 14/1, pp. 93-106, [http://dx.doi.org/10.1016/0886-1633\(93\)90046-R](http://dx.doi.org/10.1016/0886-1633(93)90046-R). [40]

Chapter 2. Making Policy

Evaluation happen: What are the institutional underpinnings?

Sound institutional set-ups can provide guidance and incentives to conduct evaluations across government in a systematic way. They can create conditions for transparency and accountability in the management of evaluations, and help promote the use of results in policy-making. This chapter presents the main aspects of the different institutional set-ups used by countries to promote evaluation as a practice within government. The chapter introduces countries' legal and policy frameworks for evaluation and discusses the nature of policy evaluation guidelines. The chapter identifies the key institutional actors in charge of the management of evaluation within the executive, such as centres of government, ministries of finance and autonomous agencies, and underlines the role of supreme audit institutions beyond the executive. Finally, the chapter stresses the importance of co-ordination mechanisms to enable greater alignment and sharing of practices across institutions.

Key Findings

- **Institutionalisation matters for effective implementation and use of evaluation.** It can provide useful incentives to ensure that evaluations can be conducted, as well as improve their quality through internal management and control tools.
- **Legal frameworks constitute the key basis** to embedding the practice of evaluations across government in a systematic way. Around two-thirds of responding countries have created a legal basis for requiring and enabling policy evaluation.
- **Policy frameworks** have also been developed in about half of the countries surveyed. These frameworks can give strategic direction for a specific sector or thematic area and can help to support the implementation of quality evaluation. They also have the potential to provide high-level guidance and clarity.
- **The centre of government is the main actor that provides strategic direction for policy evaluation.** Survey responses show that the CoG plays a crucial role in embedding a whole-of-government approach to policy evaluation and often has the widest mandates. It also provides incentives for other institutions to use evaluation findings.
- **Ministries of finance also play a very important** role and have responsibilities in many countries. Policy evaluation can help in enhancing the quality of public expenditures and in delivering improved results and performance. .
- **Ministries of planning play a significant role in about a sixth of the sample.** This trend can be explained by the relevance of national development plans in Latin America countries, and the mandate given to Ministries of Planning to evaluate these strategic plans.
- In some countries, **autonomous agencies** have also taken up competences related to policy evaluation across government. They are well placed to conduct independent, transparent and accountable evaluations. Outside the executive, **Supreme Audit Institutions** often play an important role.
- While institutions both within and outside the executive play key roles in establishing evaluation practices, **a truly embedded evaluation system benefits from coordination mechanisms**, which can provide enable greater alignment and sharing of practices across institutions.
- **There are trade-offs between ensuring the independence of evaluation, and increasing its influence when choosing where to locate the responsibility for evaluation:** locating the responsibility close to political decision-making power may prove effective to commission evaluations and to follow up on commitments by different ministries. However, agencies endowed with technical autonomy may yield a perception of transparency, unbiased judgement and accountability, which is conducive to greater trust in the results.

Understanding the institutional set-up for policy evaluation

Why does institutionalisation of evaluation matter?

A sound institutional set-up can contribute to aligning isolated and unplanned programme evaluation efforts into more formal and systematic approaches, with the ability to prioritise and to set standards for methodologies and practices (Gaarder and Briceño, 2010^[55]).

An institutional set-up can provide incentives to ensure that evaluations are effectively conducted. For instance, Mackay (2007^[58]) describes such incentives as carrots (positive encouragement and rewards for conducting policy evaluation and utilising the findings); sticks (penalties for institutions or individual civil servants who fail to take performance and policy evaluation seriously); and sermons (include high-level statements of endorsement and advocacy concerning the importance of evaluations).

Through the use and promotion of internal management and control tools in governmental institutions, a sound institutional set-up has the potential to promote the principles of transparency and accountability in the management of evaluations (Gaarder and Briceño, 2010^[55]) (Parkhurst, 2017^[32]). Thus, it can contribute to protecting policy evaluation practices from undue political influence and from undermining bureaucratic practices. This is critical as policy evaluation is a key component to ensure accountability (Lázaro, 2015^[54]). Finally, an institutional set-up can contribute to improving the comparability and consistency of results across time, institutions, and disciplines; allowing the continuity of data interpretation.

Nevertheless, laws or decrees, in and of themselves, do not ensure the effectiveness of a policy evaluation system. In some contexts, rigid institutional set-ups can even have adverse effects. For example, formal measures to undertake and use evaluation can create a fear of sanctions, which can prevent risk-taking, experimentation and innovation in policy and programme design (OECD, 2019^[59]; Brown and Osborne, 2013^[60]; Flemig, Osborne and Kinder, 2016^[61]). Additionally, an excessively rigorous system may turn public institutions into formalistic bureaucracies (Schillemans and Bovens, 2011^[62]). Hence, instead of being perceived as a learning tool, evaluations run the risk of legitimatising or reinforcing prevailing power structures (OECD, 2019^[59]).

Despite these limitations, the existence of a sound legal framework can be an important measure to promote policy evaluation and to clarify institutional responsibilities from a legal perspective. There is not a single recipe to institutionalising policy evaluation across government. Policy evaluation is characterised by a high diversity of institutional approaches across countries. According to Jacob et al (2015^[63]) “few normative claims exist regarding how evaluation should be embedded in the architecture of governance”. Factors such as the political system, public administration cultures, and the rationale for evaluation, shape the development and characteristics of evaluation cultures.

The subsequent sections intend to map and identify the main institutionalisation trends, including legal and policy frameworks and key actors. The chapter will focus on two main dimensions: the existing legal and policy framework and its key features, which provide the legal basis to undertake policy evaluations and the macro-level guidance on when and how to carry out those; and the identification of institutional actors with allocated resources and mandates to oversee or carry out evaluations.

While institutional set ups differ, the analysis mostly focuses on institutions with responsibilities within the executive branch, even if it also considers the role of supreme audit institutions. Subnational governments, parliaments and civil society are certainly influential to institutionalise policy evaluation and critical in facilitating demand for evaluation. Therefore, the fact that the report mainly draws on data concerning the executive branch calls for caution before making any inferences for other institutions outside of the scope of the survey.

What does the institutionalisation of evaluation mean?

For the purpose of this report, institutionalisation is defined as the systematic process of embedding evaluation practices into more formal and systematic approaches (Gaarder and Briceño, 2010^[55]; Lázaro, 2015^[5]). Both exogenous and endogenous factors can trigger such a process. Studies have highlighted a variety of factors for the institutionalisation of policy evaluation, such as the existence of a democratic system with a vibrant and vocal opposition, or the presence of influential evaluation champions - such as Congress, the presidency or the Minister of Finance - to lead the process (Gaarder and Briceño, 2010^[55]). Often, policy evaluation systems can be modified following crises: these crises can be financial, fiscal or result from major disasters and pandemics. The policy response requires some restructuring of the public administration, or concerns public and health safety or a general lack of performance in a policy domain, thus calling for better understanding of what works in this policy area.

In countries with longstanding traditions for evaluation, such as Australia, Canada, the United States, and the United Kingdom, government-wide evaluation cultures were initially developed with a main focus on improving performance of public expenditure and were related to the diffusion of performance budgeting, as was first initiated in the US (Jacob, Speer and Furubo, 2015^[63]). In many European countries, the growth of the welfare state, particularly in the 1970s and 1980s, in a context of slow growth and sluggish resources, spurred significant demands for policy evaluation. In others, such as Mexico, the need to have a better understanding of the impact of sectoral policies seemed to have worked as the primary driving force, in addition to the widely recognised need to better evaluate poverty, which was part of the impetus for setting up CONEVAL, the Mexican agency for policy evaluation.

In addition, international organisations and development banks have played an important role for the development of evaluation systems. European Union (EU) membership and EU Structural Funds, for instance, seem to have been crucial for the dissemination and promotion of policy evaluation in some European countries, given the strict accountability requirements related to the use of these funds (Olejniczak, Raimondo and Kupiec, 2016^[64]).

Increasing demand by citizens for more openness, transparency and better services, and the necessity to improve public sector performance have been identified as additional factors for the development of policy evaluation systems in the recent past in countries without longstanding experience.

The institutional set-up can adopt many shapes and levels of robustness. While in some countries policy evaluations are promoted through a whole-of-government legal and/or policy framework and a central institution with responsibilities across government, in others policy evaluations are the sole responsibility of line ministries with more fragmented approaches.

What has impeded institutionalisation so far?

Although institutionalisation is critical for building a policy evaluation system, countries face major challenges relating to (1) the establishment of a whole-of-government strategy, (2) human resources (capacity and capabilities) (3) political interest and demand for policy evaluation (4) financial resources required for policy evaluations and (5) the availability and quality of data.

According to the survey, the greatest challenge that countries encounter in promoting evaluation across government is the absence of a **strategy that promotes a whole-of-government approach to policy evaluation** (Figure 1.8 in Chapter 1). The institutionalisation process involves a wide variety of actors, many of them subject to inertia or resistance to change. Without effective guidance – for instance on mandates, timing and resources – public organisations may fail to make coordinated decisions and agree on a common vision, mission and shared goals, which are all necessary steps in setting up a policy evaluation system (Cinar, Trott and Simms, 2018^[65]). Successful institutionalisation can also benefit from the engagement of external stakeholders such as citizens and academia, whose participation relies on transparency and accountability mechanisms that might be difficult to put in place without proper planning (Viñuela, Ortega and Gomes, 2015^[66]).

A second factor identified by countries is linked to **human resources, in terms of capacity and capabilities** for policy evaluation. This is partly related to the fact that civil servants lack the time to absorb the new practices associated to evaluation, especially when those are not directly related to their operational priorities (Cinar, Trott and Simms, 2018^[65]), (Bossuyt, Shaxson and Datta, 2014^[67]).

A third major challenge perceived by respondents relates to the **low political interest in, and demand for, policy evaluation**. Commitment at the highest political level is a key enabler to successful governance reforms (OECD, 2018^[2]). Without strong political interest and demand for policy evaluation, it is difficult to find incentives for civil servants usually busy in managing day-to-day responsibilities. It becomes also challenging for knowledge brokers and for the advocates and transmitters of evaluations to engage with policymakers and civil servants in tasks outside their immediate area of responsibility.

(Liverani, Hawkins and Parkhurst, 2013^[68]). The demand for evaluation may be caught in a vicious circle, where the lack of demand comes from insufficient understanding of evaluation practices and purpose, which itself comes from a lack of experience with evaluations, due to weak demand for Mackay (2007^[58]).

Another challenge is the **financial resources**. Institutionalising policy evaluations can be financially and labour intensive. While the production of consistent data and the dissemination of results play a crucial supporting role in the institutionalisation process (Maeda, Harrit and Mabuchi, 2012^[69]), this demands capacity for consistent estimation methods, communication facilities, and time (Zida et al., 2017^[70]).

Finally, the **limited availability and quality of data** across government agencies and departments can also be a major challenge. Data is a strategic asset to improve policy design, service delivery and the operations of the machinery of government. Nonetheless, enabling the strategic use and quality of data requires human and technical capabilities, especially the willingness of the public servants to use data, as well as an investment in data analytical tools (van Ooijen, Ubaldi and Welby, 2019^[71]).

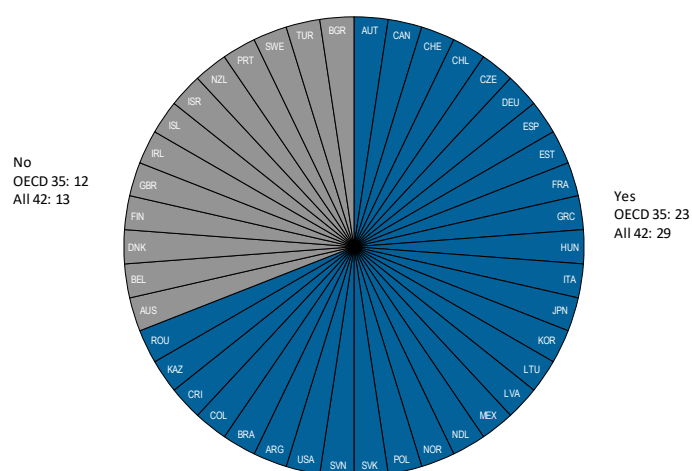
Anchoring policy evaluation in legal & policy frameworks

An adequate legal and policy framework constitutes a solid basis to embed the practice of evaluations across government in a systematic way. However, there is no one-size-fits-all: countries have developed laws, policies and guidelines to promote evaluations in various ways.

Ensuring solid legal frameworks for policy evaluation

A majority of countries (29 countries, 23 OECD countries) have developed a legal framework that guides policy evaluation, as is the case in Chile, Norway, Poland, and Slovenia. The fact that over two-thirds of responding countries have created a legal basis for policy evaluation underlines the importance that OECD member and partner countries attribute to this practice across government (Figure 2.1)

Figure 2.1. Availability of a legal framework guiding policy evaluation across government



Note: n=42 (35 OECD member countries). Answers reflect responses to the question, "Is there a legal framework guiding policy evaluation across government?"

Source: OECD Survey on Policy Evaluation (2018).

The legal anchors of evaluation can vary substantially across countries. Some countries have specific stipulations in their constitutions while others focus on primary or secondary laws. An overview of the nature of the legal framework is provided below Table 2.1.

Table 2.1. Nature of legal framework for policy evaluation

	Constitution	Primary legislation (laws or equivalent)	Secondary/subordinate legislation
Austria	○	●	○
Canada	○	●	○
Chile	○	●	○
Czech Republic	○	●	○
Germany	●	●	●
Estonia	○	●	●
France	●	●	●
Greece	○	●	●
Hungary	○	●	●
Italy	○	●	●
Japan	○	●	●
Korea	○	●	●
Latvia	○	●	●
Lithuania	○	●	●
Mexico	●	●	●
Netherlands	○	●	●
Norway	○	○	●
Poland	○	●	●
Slovakia	○	○	●
Slovenia	○	●	○
Spain	○	●	●
Switzerland	●	●	●
USA	○	●	○
OECD Total			
● Yes	4 (17%)	21 (91%)	17 (74%)
○ No	19 (83%)	2 (9%)	6 (26%)
Argentina	○	○	●
Brazil	○	○	●
Colombia	●	○	○
Costa Rica	●	●	●
Kazakhstan	○	○	●
Romania	○	○	●

Note: n=29 (23 OECD member countries). 13 countries (12 OECD members) are not included as they answered that they do not have a legal framework guiding policy evaluation across the government. Answers reflect responses to the question, “At what level is policy evaluation across government legally embedded? (Check all that apply)”.

Source: OECD Survey on Policy Evaluation.

Constitutional provisions for policy evaluation

Requirements for policy evaluation can be incorporated at the level of the constitution. This reflects a significant commitment and provides an important mandate to the government in this area. Moreover, it institutes policy evaluation as a long-term policy, as the incorporation of elements in a constitution reflects a great degree of consensus among different political actors, which usually goes beyond electoral mandates. Germany, France, Mexico, Switzerland, Colombia, and Costa Rica have specific provisions within their constitution relating to policy evaluation (Box 2.1).

Box 2.1. Examples of policy evaluation-related principles found in national constitutions

The Constitution of **Switzerland** requires the Federal Assembly to ensure the evaluation of federal measures in terms of their effectiveness.

The **German** Constitution states the necessity to conduct evaluations of financial assistance grants on a regular basis (Article 104b).

In **France**, Articles 47-2 of the Constitution of the 5th Republic mandate the French Supreme Audit Institution (Cour des Comptes) to assist the Government and Parliament in the evaluation of public policies, among other duties (See Box 2.2 on France's embedded policy evaluation framework).

Moreover, the **Mexican** Constitution's Article 134 requires that economic resources be managed and used efficiently, effectively and transparently, and that the results of such use be assessed by technical agencies, in order to guarantee an optimal budget allocation.

Colombia's Constitution contains a larger number of articles that establish evaluative activities, such as the prescription of the national planning entity to organise the evaluation of public administrations' management and performance (Article 343).

Lastly, the **Costa Rican** Constitution's Article 11 prescribes the evaluation of the results and accountability of all public institutions as well as the fulfilment of civil servants' duties.

Source: OECD (2018) Survey on Policy Evaluation.

As Box 2.1 shows, constitutional provisions can give responsibilities to particular entities and can define approaches and scopes of evaluation practices. Constitutional provisions might differ in terms of who they mandate to conduct evaluations. For instance, the French constitution mandates the supreme audit institution to assist the government and the parliament in policy evaluation. In Mexico, the constitution states that technical agencies should evaluate the use of national resources, while in Colombia, the national planning entity is required to organise evaluations.

Constitutional mandates may also have specific provisions regarding the scope and object of evaluation. For example, the Swiss Constitution requires federal measures to be evaluated in terms of their effectiveness, whereas the French Constitution requires legislative proposals to be evaluated in terms of their impact. Even more specifically, the German Constitution mandates the regular evaluation of financial assistance grants, while the Colombian one requires the evaluation of public administrations' management and performance. Constitutional provisions can also focus on the particular duties that civil servants have, as is the case in Costa Rica.

These provisions can largely shape the configuration of a country's evaluation system. While in countries such as Colombia, the evaluation system is linked to development planning, the German system is closely oriented to spending reviews and the role of the Parliament in assessing the Federal Government's performance. The main elements of the French legal framework are presented below (Box 2.2).

Box 2.2. France's embedded policy evaluation framework

France implemented a legal framework for policy evaluation embedded at three different levels: the constitution, primary legislation and secondary legislation.

At the **constitutional** level, article 47-2 mandates the French Supreme Audit Institution (Cour des Comptes) to assist the parliament and the government in evaluating public policies. The results are made available to government and citizens through publication of the evaluations. Evaluative activities are also expressed in articles 39 and 48 of the Constitution.

In terms of **primary legislation**, articles 8, 11 and 12 of the organic law number 2009-403 on the application of article 34-1 of the Constitution requires legislative proposals to be subject to *ex ante* impact assessment. Assessment results are then annexed to the legislative proposal as soon as they are sent to the Supreme Administrative Court (Conseil d'État).

On the **secondary legislation** level, article 8 of the Decree No. 2015-510 states that all legal draft proposals affecting the missions and organisation of decentralised State services should be subject to an impact assessment. The main objective is to check the alignment between the objectives pursued by the proposal and the resources allocated to decentralised services.

Additionally, France has a number of **circulars from the prime minister** that relate to evaluation. On October 12th 2015, the circular related to the evaluation of norms, and in May 2016 to the impact evaluation of new law projects and regulatory texts.

Source: OECD (2018) Survey on Policy Evaluation, Constitution de la V^e République, and the respective articles from Legifrance (<https://www.legifrance.gouv.fr>).

Primary and secondary legislation on policy evaluation

Primary laws or equivalent and secondary legislation (decrees, ministerial resolutions or equivalent) represent the most frequent legal basis for institutionalisation.

Primary legislation frameworks differ substantively across countries. Some countries have framed evaluation as part of larger public management laws. This is the case of the United States, with the Foundations for Evidence-Based Policymaking Act of 2018 (see Box 2.3). In the field of policy evaluation, the act mandates the Office of Management and Budget (OMB) to develop guidance and advice in policy evaluation. The law includes a provision that requires agencies to submit annual evaluation plans, which shall "describe key questions for each significant evaluation study that the agency plans to begin in the next fiscal year". It also mandates government's agencies to: (1) designate a senior employee as evaluation officer to coordinate evidence-building activities; (2) develop capacity assessments which "shall contain an assessment of the coverage, quality, methods, effectiveness, and independence of the statistics, evaluation, research, and analysis efforts of the agency"; (3) implement OMB guidance for programme evaluation; (4) identify "key skills and competencies, establish or update an occupational series, and establish a new career path" on programme evaluation (115th Congress, 2019^[7]).

Box 2.3. Building the institutional foundations for evidence-based policymaking in the US

The Foundations for Evidence-Based Policymaking Act of 2018, which resulted from the work of a Bipartisan Commission of Congress, was signed and enacted into law on January 14th, 2019. The Evidence Act aims for federal agencies to better acquire, access, and use evidence to inform decision-making. It includes three Titles and has a significant impact in terms of the institutionalisation of evidence across federal government:

1. Federal Evidence-Building Activities
2. Open Government Data Act
3. Confidential Information Protection and Statistical Efficiency (CIPSEA)

Accordingly, the Act mandates **evidence-generating activities across agencies**, open government data, confidential information protection, as well as skills and capacity building. This Act matters in that it elevates programme evaluation as a key agency function, calling on agencies to strategically and methodically build evidence in a coordinated manner.

The **implementation approach** of this Act is phased and coordinated. Its first and foundational phase (“Learning Agendas, Personnel and Planning”) centres on developing learning agendas, identifying relevant personnel, their roles and responsibilities, and undertaking planning activities. The purpose of the **learning agendas** is to promote deliberate and strategic planning of evidence-building activities. In creating the learning agendas, agencies are required to identify and set priorities for evidence building, in consultation with various stakeholders. The second element – **personnel** – involves three newly designated positions (**Chief Data Officer, Evaluation Officer, and Statistical Official**), who spearhead activities pertaining to Phase 1 of the implementation, including reporting requirements. These individuals also serve on a **Data Governance Body** inside of their respective agency, which is concerned with managing data as a strategic asset to fulfil the agency’s mission as well as addressing the priorities identified in the agency’s learning agenda.

The last element of the first phase of implementation consists of various planning activities. This includes developing **annual evaluation plans**, which outline the specific evaluations that each agency intends to carry out to address its learning agenda priorities. Furthermore, agencies are required to undertake **capacity assessments** in order to assess their ability to carry out evidence-building activities like performance measurement, fact-finding, etc. Finally, agencies are also required to identify data to answer the questions outlined in their learning agendas.

The learning agenda activity is intended to drive all other evidence-building activities. Other aspects of the implementation of the Act include: “Open Data Access & Management”, “Data Access for Statistical Purposes”, and finally “Programme Evaluation”. The fourth phase of implementation will consist of the Office of Management and Budget (OMB) issuing guidance on Programme Evaluation Standards and Best Practices as well as on Evaluation Skills and Competencies (with the Office of Personnel Management).

Sources: (The Statistical Reform Promotion Council, 2017^[72]), (The Committee on Promoting EBPM, 2017^[73]), (The Cabinet Secretariat, 2019^[74]), (United States Office of Management and Budget, 2019^[75])

Other countries have issued specific legislations on policy evaluation, such as Japan with the Government Policy Evaluations Act (see Box 2.4) and Korea with the Framework Act on Government Performance Evaluation (See Box 2.5).

Box 2.4. Institutionalisation of Policy Evaluation in Japan

To provide the policy evaluation system with a clear-cut framework and improve its effectiveness, Japan has enacted the Government Policy Evaluations Act of 2001, which provides an overarching framework for the policy evaluation and clarifies the role of each ministry in the evaluation of policies. It requires appropriate implementation of policy evaluations prior to the adoption of policies, and specifies how policy evaluations conducted by the Ministry of Internal Affairs and Communications (MIC) should be conducted.

Under the act, the “Basic Guidelines for Implementing Policy Evaluation” offer guidelines for the development of basic plans by for each ministry to develop an evaluation plan in order to promote a whole-of-government approach to evaluation. The ministries’ “Basic Plan for Policy Evaluation”, cover a period of 3 to 5-year and incorporate policy evaluation into public management cycle such as “Plan, Do, Check, Act (PDCA)”.

The MIC has also developed various guidelines to support use and publication of evaluations (e.g., “Policy Evaluation Implementation guidelines” (2005,) and “Guidelines for Publication of Information on Policy Evaluation” (2010)).

The Administrative Evaluation Bureau (AEB)

The AEB formulates standard rules and guidelines for conducting policy evaluations, aggregates all policy evaluation reports across the government, and conducts reviews to improve the quality of those evaluations. In 2012, the AEB introduced a standard format across ministries for *ex-post* evaluation of major policy that made it easier to read and compare the evaluation reports. Besides, the AEB set up Portal Site for Policy Evaluation in FY 2012, which provides links to policy evaluation data including analysis sheets and evaluation reports publicised by each ministry to ensure transparency and accountability.

Sources: (Ministry of Internal Affairs and Communications, 2017^[76]), (The Ministry of Internal Affairs and Communications, 2010^[77]). (The Ministry of Internal Affairs and Communication, 2005^[78])

Box 2.5. Korea's Policy Evaluation System

The 100 Policy Tasks of the Republic of Korea

The Republic of Korea's "Five-Year Plan", set by the State Affairs Planning Advisor Committee, consists of an overarching vision, policy goals and strategies, and the 100 Policy Tasks. The tasks were selected through a review of over 200 pledges and nearly 900 breakdowns of pledges proposed to, and received from, the public during the presidential election campaign. The 100 policy tasks include a system for comprehensive monitoring and management, conducted with close cooperation between the Presidential Commission on Policy Planning, Office for Government Policy Coordination, and the Government Performance Evaluation Framework. Government performance evaluation implementation plans are released annually, providing overviews on evaluation structure and more detailed evaluation plans organised by levels of governance, from central administrative agencies to local governments.

Framework Act on Government Performance Evaluation

The Government of Korea established a performance management system by enacting the Framework Act on Governance Performance Evaluation (FAGPE) in 2006. This law aims to improve the efficiency, effectiveness, and accountability of the government administration by establishing the fundamental principles, institutional foundation, management strategies and execution plans on which government performance evaluations can be implemented (Roh, 2018^[79]).

Prior to 2006, the performance management and evaluation systems of the Korean government were dispersed and consisted of different programs under various agencies (Yang and Torneo, 2016^[80]). The act aimed to improve and integrate the different performance management systems of all government organisations. This practice has enabled systematically managing their performance under the Government Performance Evaluation Committee (GPEC). This Committee oversees all government performance management and evaluation systems, and provides consistency and stability in government performance management (Roh, 2018^[79]).

Sources: OECD (2018) Survey on Policy Evaluation, (Roh, 2018^[79]), (Yang and Torneo, 2016^[80]).

Finally, some countries have regulated policy evaluation as part of their budgetary governance framework, as is in the case of Estonia, Germany and Italy, among others. In the case of Germany, the Federal Budget Code of 19 August 1969, as last amended in 2017, provides the guidelines for presenting and authorising the annual budget. This code includes procedures of the budget implementation, auditing and evaluation (OECD, 2014^[81]). The Federal Budget Code is mandatory to the Federal Parliament, the Federal Government and individual federal ministries.

Secondary legislation can also support and streamline constitutional mandates or primary legislation. This type of legislation usually provides more detailed policy frameworks on evaluation; usually containing specific information concerning evaluation annual plans, timing, selection criteria, etc. This is the case of the Netherlands, with the Regulation of the Minister of Finance (15-03-2018), which lays down rules for periodic evaluation based on the Accountability Act of 2016. In these arrangements, the Ministry of Finance establishes the scope, actors involved, timing, and minimal quality criteria requested in an evaluation.

Some countries frame their evaluation system only through regulations and acts issued by the executive. This is the case of Norway, Slovakia, Argentina, Brazil, Kazakhstan and Romania. The cases of Norway and Argentina illustrate this situation:

- Norway, within the Ministry of Finance's Regulation for Financial Management in Central Government on Financial Management in Central Government (2003), mandates all agencies to conduct evaluations to gather information on efficiency, objective achievement and results, in all or some of their areas of responsibility and activities.
- Argentina, with the decree 292/2018 on monitoring and evaluation guidelines lays down two main objectives: (1) to mandate the National Council for the Coordination of Social Policies as the body responsible for preparing and executing the Annual Policy and Social Programmes Monitoring and Evaluation Plan; and (2) to provide technical assistance to ministries and national organisations for the evaluation and monitoring of policies, programs, plans and projects with social impact.

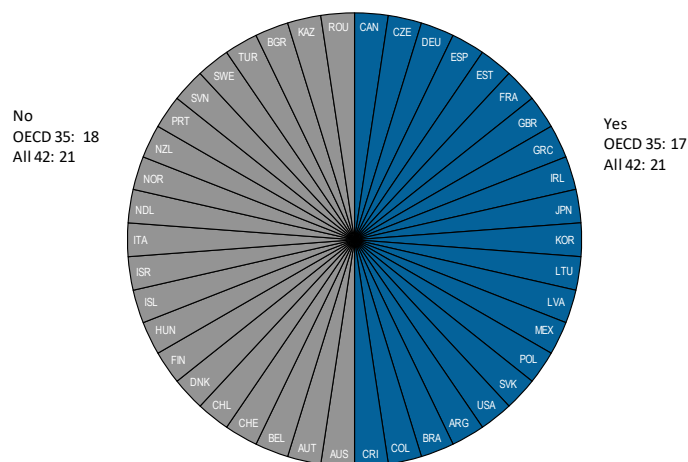
Creating a supportive policy framework

A policy framework is generally a document or set of documents that provides strategic direction, guiding principles and courses of action to the government for a specific sector or thematic area. Policy frameworks can include different legislative acts, but this is not necessarily always the case, and some ministries can internally adopt evaluation policies in the form of guidelines without issuing any specific regulation. In any case, a clear policy framework can help to:

- conduct different aspects of policy analysis in a credible and rigorous manner, which supports the implementation of quality evaluation.
- provide high-level guidance and clarity for institutions by outlining overarching best practices and goals, generally taking the form of an institution-wide guidance document that describes implementation or standards for policies, establishes hierarchies and categories, and outlines the exigence or rationale behind stated goals.

Half of surveyed countries (21 in total, including 17 OECD countries) developed a policy framework for organising policy evaluation across government (Figure 2.2). Among those, a number of countries implemented both a legal and a policy framework (19 in total, including 15 OECD countries). Countries include Estonia, Japan, Korea, Colombia and Costa Rica.

Figure 2.2. Availability of a policy document on policy evaluation across government



Note: n=42 (35 OECD member countries). Answers reflect responses to the question, "Apart/under the umbrella of a legal framework, has your government developed a policy framework for organising policy evaluation across government?"

Source: OECD Survey on Policy Evaluation (2018).

The Czech Republic established a “Methodological guidance for Evaluation in the 2014-2020 programming period”. Taken with its legal framework that has more of an operational focus, this methodological guidance contributes to creating a comprehensive evaluation framework. Korea’s Framework Act on Governance Performance Evaluation (FAGPE) is complemented with two policy documents: a Basic Plan of the Government Performance Evaluation (2017-2019), and Operational Rules of the Government Performance Evaluation Committee.

Policy frameworks tend to allocate institutional responsibilities for evaluation, with a total of 17 countries surveyed (of which 14 are OECD) outlining this in their evaluation policy (see Table 2.2). In Germany for instance, the Instructions for Economic Efficiency Investigations (Arbeitsanleitung Einführung in Wirtschaftlichkeitsuntersuchungen) intend to guide the realisation of economic evaluations.

Table 2.2. Features of the framework for policy evaluation

	Objectives or expected results of the evaluation policy	Policy areas (thematic) or programmes covered by the evaluation policy	Responsibilities of government institutions concerning policy evaluation	Requirement for government institutions to undertake regular evaluation of their policies	Standards for ethical conduct	Requirements related to the quality standards of evaluations	Requirements related to stakeholder engagement	Requirements related to evaluation reporting	Requirements related to the use of evaluation findings into policy planning making
Canada	●	○	●	●	○	○	○	●	○
Czech Republic	○	○	○	○	●	○	●	●	●
Germany	●	●	●	●	●	●	●	●	●
Estonia	●	○	●	○	●	○	●	●	○
France	●	○	●	●	○	●	○	○	○
Great Britain	●	○	●	●	●	○	○	●	●
Greece	●	●	●	●	○	●	●	●	●
Japan	●	●	●	●	○	●	○	●	●
Korea	●	●	●	●	●	●	●	●	●
Latvia	●	●	○	●	○	○	●	●	○
Lithuania	○	●	●	○	●	○	●	○	○
Mexico	●	●	●	●	●	●	●	●	●
Poland	○	●	●	●	●	●	○	●	●
Slovakia	●	●	●	○	○	●	○	○	●
Spain	●	●	●	●	●	●	○	○	○
USA	●	○	●	●	○	○	●	○	○
OECD Total									
● Yes	13 (81%)	10 (63%)	14 (88%)	12 (75%)	9 (56%)	9 (56%)	9 (56%)	11 (69%)	9 (56%)
○ No	3 (19%)	6 (37%)	2 (12%)	4 (25%)	7 (44%)	7 (44%)	7 (44%)	5 (31%)	7 (44%)
Argentina	●	●	●	●	○	○	●	●	●
Brazil	●	○	○	●	○	○	●	●	○
Colombia	●	○	●	●	○	●	●	●	●
Costa Rica	●	●	●	●	○	●	●	●	●

Note: OECD Survey on Policy Evaluation (2018).

Source: n=20 (16 OECD member countries). 21 countries (18 OECD member countries) are not included as they answered that they do not have a policy framework for organizing policy evaluation across government. Data is not available for Ireland. Answers reflect responses to the question, “Which elements do(es) the document/s referred to under Q4 and Q5 cover concerning policy evaluation across government? (Check all that apply)”. The documents referred to under Q4 and Q5 are the ones stipulating a policy framework organising policy evaluation across government. The option “Other” is not included.

Evaluation plans, or requirements for government institutions to undertake regular evaluation of their policies are also common among respondents (16 overall, and 12 OECD countries). Such evaluation plans exist in Spain, where an Action Plan for the evaluation of the Spending Review, an Annual Plan of Normative Impact, and a Master Plan of Spanish Cooperation for 2018-2021 have been created. In Mexico as well, an Evaluation Programme is published every year since 2007.

Interestingly, some countries who lack an overarching legal framework nevertheless created a policy framework to promote evaluation. These include Canada, the United Kingdom and Ireland. The United Kingdom's Treasury has a set of policy frameworks that give government guidance on evaluation, such as the Green book, which is particularly focused on centre of government's¹ responsibilities concerning evaluation. The other guide, the Magenta Book, is to be used by policy analysts, policy makers in all levels of government, including central and local, and the voluntary sector, to institutionalise good evaluation practices. Another key reference is the European Commission's guidance, which presents key principles for the implementation and design of evaluations (Innovate UK, 2018^[82]).

Many policy frameworks state objectives or expected results (17 surveyed countries overall including 13 OECD ones). Canada for instance, which does not have a legal framework for policy evaluation, has a Policy on Results (See Box 2.6). One of its objectives is to improve the achievement of results across government, and it expects federal departments to measure and evaluate their performance to ultimately improve policies, programmes and services (Canada Treasury Board, 2016^[81]).

Box 2.6. Canada's Policy on Results

In July 2016, the Government of Canada launched a Policy on Results, which seeks to improve the achievement of results across government and better understand the desired results and the resources used to achieve them.

The responsibility for the implementation of this policy mainly falls under the Treasury Board. This body is responsible for promoting the use of evaluation findings into policymaking and defining and updating the evaluation policy.

The policy establishes that all government departments should have an evaluation unit. On the other hand, line ministries are responsible for establishing a departmental results framework. For the implementation of the policy, the Treasury Board of Canada has, among others, the following competences:

- It can require departments to undertake specific evaluations and participate in centrally-led evaluations;
- It can initiate or undertake resource alignment reviews;
- It approves line ministries departmental results frameworks and any changes to their organisations' core responsibilities.

This policy complements Canada's Financial Administration Act, which requires the evaluation of grants and contributions programs every five years.

Source: (Canada Treasury Board, 2016^[81]); <https://laws-lois.justice.gc.ca/PDF/F-11.pdf>.

Policy frameworks may focus on particular policy areas and programmes, as 12 countries surveyed - including 10 OECD countries- have implemented thematic or specific policy frameworks. Ireland for example, does not have a formal legal framework for policy evaluation, and instead has guidelines for Regulatory Impact Analysis and a public spending code.

In addition to institutionalising responsibilities, policy frameworks may also include provisions regarding the quality and use of evaluation, such as standards for the ethical conduct of evaluators, requirements for stakeholder engagement, reporting, and use of findings into policy-making. Country practices concerning the promotion of quality and use of evaluation in policy frameworks are discussed in Chapter 3. Germany, Korea, Greece and Costa Rica have comprehensive policy frameworks that include these elements on the quality and use. Greece's Manual of Inter-Ministerial Coordination includes all elements in the table above, but for standards for ethical conduct. Another example is Costa Rica's National Evaluation Policy (Box 2.7).

Box 2.7. The National Evaluation Policy (PNE) in Costa Rica

The National Evaluation Policy (PNE) was established by the Ministry of National Planning and Economic Policy (Mideplan), in coordination with line ministries, academics and civil society. This Policy serves as an instrument to establish a framework to strengthen the progress of evaluations in the public sector.

The PNE particularly aims at improving public management by promoting evaluation as an instrument for decision-making, learning, control of public resources, and accountability. The PNE focuses on four axes of action:

1. Evaluation in the Management cycle for Development Results: this aims to increase the evaluability conditions in public interventions through a joint work between Mideplan and the Ministry of Finance (e.g. through technical and methodological guidelines for the use of evaluations in social programs).
2. Institutionalisation of an evaluation framework: this aims to improve the design and management of public interventions based on evidence (e.g. public repository with previews of evaluations in the public sector).
3. Capacity building in evaluation: this aims to increase the quality of evaluations made in the public sector (e.g. training to civil servants on the design and implementation of evaluation).
4. Stakeholders' participation: this aims to increase the participation of stakeholders in the evaluation process (e.g. spaces for dialogue and interaction between different actors within government and external ones such as civil society organisations).

Source: (Mideplan, 2018^[83]).

The role of guidelines

Guidelines and other supporting documents such as White Books on evaluation can assist policy makers in conducting policy evaluation successfully. Evidence shows that the majority of countries (31 countries and 26 OECD members) has guidelines to support the implementation of policy evaluation across government (See Table 2.4). Such guidelines for policy evaluation generally intend to assist all those participating in the implementation of a policy in better planning, commissioning and managing its evaluation (OECD-DAC, 2009^[84]). Effective implementation requires a structured approach about how the policy will deliver a service and program successfully, considering risks and implementation issues (Department of the Prime Minister and Cabinet (Australia), 2013^[85]). Countries such as Australia, Finland, New Zealand and Portugal only have guidelines but do not report a policy or a legal framework.

Guidelines mostly refer to the reporting of evaluation results, followed by the identification and design of evaluation approaches, quality standards of evaluations, and use of evaluation evidence (See Chapter 3). Around half of countries' toolkits refer to the design of data collection methods,

independence of evaluations, and stakeholder engagement in the evaluation process. Canada presents a significant number of guidelines for the implementation and evaluation of policies (see Box 2.8) and the United States recently updated and consolidated its guidance on programme evaluation standards and practices as part of the implementation of the Foundations for Evidence Based Policy Making Act of 2018 (Box 2.3) (Office of Management and Budget, 2020^[86]).

Box 2.8. The role of frameworks and guidelines for the promotion of evidence-informed policy making (EIPM) in Canada

The Results Division of the Secretariat, successor of the Centre of Excellence for Evaluation (CEE), is responsible for evaluation activities within the Government of Canada, under the 2016 Policy on Results [See Box 2.6 on Canada's Policy on Results]. It offers useful resources, information and tools to government professionals and anyone else interested in evaluation at the federal level. Overall, the Secretariat has functional leadership regarding the implementation, use and development of evaluation practices across government. To support quality EIPM, the Results Division offers a number of useful guidelines:

- **Guide to Rapid Impact Evaluation (RIE):** this practical guide gives a range of methods for conducting RIE and advice on when and how it can be used in government. More precisely, it defines RIE, the time and resources needed to conduct one, its key benefits and challenges, and support for planning, analysis and reporting of the results.
- **Assessing Programme Resource Utilization When Evaluating Federal Programmes:** this document is made for evaluators of federal government programmes, programme and financial managers, and corporate planners. It helps them understand, plan and undertake evaluations that include the assessment of resource utilization. It provides them with methodological support to ensure that they have the knowledge and competencies to conduct quality and credible programme resource utilization assessments.
- **Theory-Based Approaches to Evaluation: Concepts and Practices:** this document introduces key concepts of theory-based approaches to evaluation and their application to federal programmes. It should be complemented by additional readings and advice for step-by-step guidance on conducting evaluations.
- **Supporting Effective Evaluations: A Guide to Developing Performance Measurement Strategies:** this guide supports departments, programme managers and heads of evaluation in developing performance measurement to support evaluation activities. It provides recommendations, tools and frameworks for conducting clear and concise performance measurement strategies as well as guidance regarding the roles of those in charge of developing such strategies.

Sources: (Treasury Board Secretariat, 2019^[87]), (Treasury Board Secretariat, 2013^[88]), (Treasury Board Secretariat, 2010^[89]).

More specifically, a majority of countries (11 countries, of which 9 OECD countries) set out guidelines for technical quality and good governance of evaluations, such as Japan (see Box 2.9). This is further explored in Chapter 3. .

Box 2.9. Basic Guidelines for Implementing Policy Evaluation by the Ministry of Internal Affairs and Communications (MIC) of Japan

With its law No. 86 of 2001, Japan enacted the **Government Policy Evaluations Act**. The act clarifies the administrative organs's obligation to evaluate policies after their adoption under a clear-cut plan, requires the appropriate implementation of policy evaluations prior to adoption, and specifies what policy evaluations should be conducted by the MIC.

For this purpose, the MIC presented the **Basic Guidelines for Implementing Policy Evaluation** to support the development of such plans by individual administrative organs and the Government's Policy Evaluation activities, in accordance to the Article 5 of the Act.

These guidelines include the purposes of both *ex-ante* and *ex-post* evaluation, different methods to measure policy impacts, recommendations regarding the use of insights from academic experts, the incorporation of evaluation results in policymaking, and the public reporting of those results.

The guidelines also indicate that the MIC shall host liaison meetings with representatives from each ministry in order to foster close communication among them and ensure the smooth and efficient implementation of a policy evaluation system.

Source: Adapted from the Ministry of Internal Affairs and Communications (2017^[76]).

The principal institutions in charge of policy evaluation and their mandates

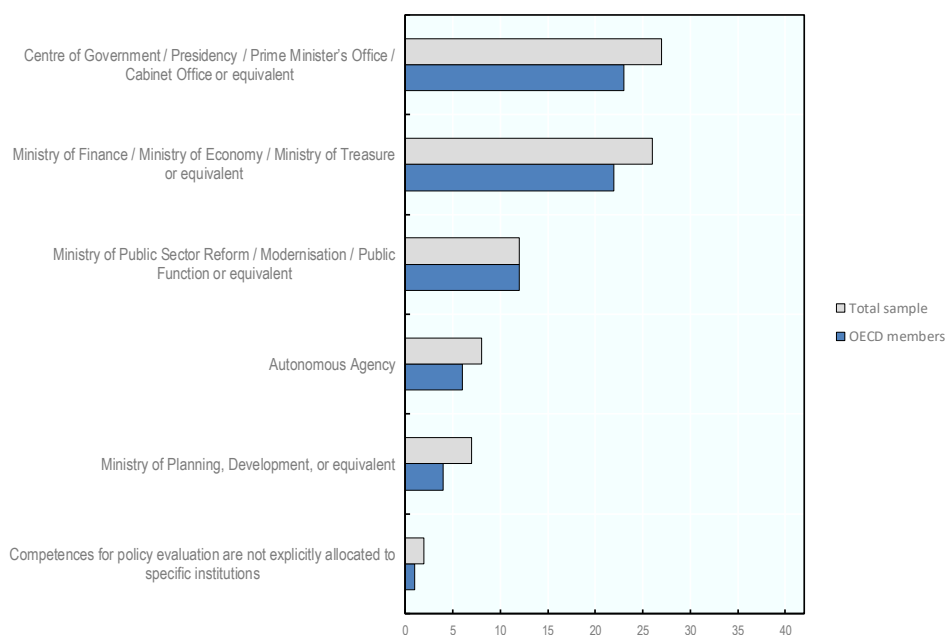
Institutions within the Executive

The survey results show that the centre of government is the principal institutions in charge of policy evaluation across government. This is the case in 27 countries, including 23 OECD countries. The second actor is the ministry of finance in 26 countries, including 22 OECD countries. Ministries of planning, development or equivalent have competences related to policy evaluation across government in 7 countries including 4 OECD countries. Ministries of public sector reform or equivalent have such competencies in 12 OECD countries.

In fact, there is often a dual role, with policy evaluation to be carried out and institutionalised in all the sectoral ministries and agencies, with some form of coordination from the centre, either through COG processes, or budget and resources related aspects with the Ministry of Finance. The central institution therefore has a key role in managing the evaluation eco-system, making sure that evaluation can take place at the right time and in the right place and that it can feed into decision making. In some cases, the core institution can also develop its own capacity for evaluation, either through the evaluation of public spending in the ministry of finance, or the evaluation of cross cutting government priorities and strategies at the centre of government.

The results shows that 40 countries have at least one institution with responsibilities related to policy evaluation across government. Consequently, a great majority of countries have chosen to allocate the mandate of coordinating policy evaluation across the executive to either one or several institutions (Figure 2.3).

Figure 2.3. Institutions within the Executive that have competences related to policy evaluation across government



Note: OECD Survey on Policy Evaluation (2018).

Source: n=42 (35 OECD member countries). Answers reflect responses to the question, "Which of the following institutions within the executive have competences related to policy evaluation across government? (Check all that apply)". Answer option "other" is not displayed.

Only in 1 OECD countries (2 countries in total), competences for policy evaluation across government are not centralised in a single institution. The fact that in the majority of OECD countries policy evaluation is conducted by more than one institution underlines the importance of steering and coordination capacities. As analysed in Box 2.10 the institutions responsible for policy evaluation across government differ across OECD countries. In most countries, the centre of government has the broader mandate and is thus well placed to conduct this horizontal task.

Box 2.10. Examples of institutions responsible for Policy Evaluation in OECD countries

Centre of government: Finland

The centre of government of Finland, which consists of the ministry of finance, the Ministry of justice and the Prime minister's office, exercises the competences related to policy evaluation. In order to enhance the use of evidence, the government established in 2014 a policy analysis unit under the Prime minister's office. The unit has the mandate to commission research projects and present evidence to support the government's decisions on future strategic and economic policy.

Ministry of finance: Chile

The Budgets Directorate (Dirección de Presupuestos), a dependent body of the ministry of finance (Ministerio de Hacienda), is the technical body in charge of ensuring the efficient allocation and use of public funds. To do so, the directorate carries out *ex ante*, impact and value-for-money evaluations of different governmental policies and programmes. Moreover, it monitors the implementation of government programmes to collect performance information, which is then introduced into the budgetary process and communicated to stakeholders.

Autonomous agency: Mexico

The National Council of Social Development Policy Evaluation (Consejo Nacional de la Política de Desarrollo Social, CONEVAL), was created in 2004 as a decentralised body with budgetary, technical and management autonomy. It has the mandate (embedded in the Constitution in 2014) to set standards, co-ordinate the evaluation exercises of the National Social Development Policy and its subsidiary actions, and provide guidelines to define, identify and measure poverty. The agency carries out or contracts out evaluations of the social policies developed by the Mexican government.

Decentralised system: Norway

Norway has a decentralised evaluation system. The Agency for Financial Management has an important role in issuing guidelines, and guiding materials. Evaluations are conducted by individual agencies, and there is a general portal managed between the Norwegian Government Agency for Financial Management (DFØ) and the national library, which contains all public evaluations from 2005 until today. Norway has moreover established a network of evaluators, chaired by the agency for financial management (EVA-Forum). Norway has a strong experience and tradition for evaluation in some sectors such as development policy or education. However, *ex post* evaluations are only carried out for certain regulations in response to requests from parliament, external groups audit office or due to legal requirements.

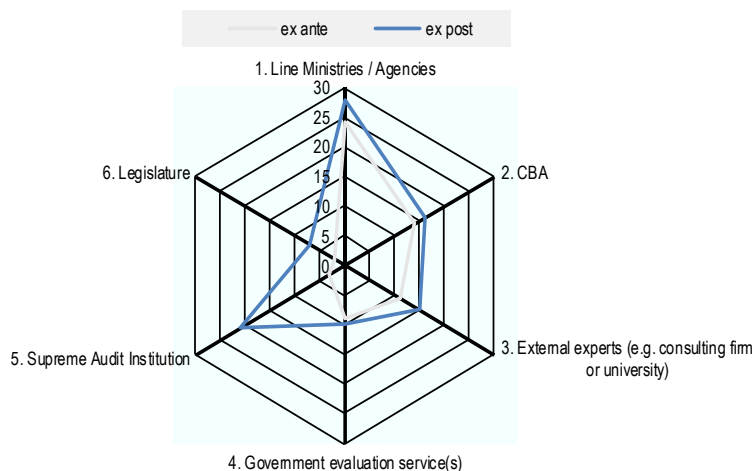
Decentralised system: Sweden

The Swedish organisational structure includes specific sectorial agencies within the executive whose main task is to perform analyses and evaluations for the government needs. In addition to the Division for Structural Policy at the ministry of finance, there are also seven sector specific evaluation agencies in Sweden², in areas such as Growth, Transport and Crime Prevention. There is also an agency for public management (Statskontoret, the Swedish Agency for Public Management), which is the Government's organisation for analyses and evaluations in all areas of state and state-funded activities.

Sources: from Government Policy Analysis Unit (2017^[90]) and Secretaria de Desarrollo Social (2015^[91]).

The degree of involvement of institutions in different types of evaluations differ. The OECD Report on Budgeting and Public Expenditures in OECD Countries 2019 finds that line ministries and agencies have a very active role in both *ex-ante* and *ex-post* evaluations (OECD, 2019^[12]). Supreme audit institutions take more of a substantial role in relation to *ex-post* reviews (See Figure 2.4).

Figure 2.4. Governance of *ex ante* and *ex post* evaluation



Note: Data for Israel and the United States are not available, Information on data for Israel: <http://dx.doi.org/10.1787/888932315602>.

Source: OECD (2018), OECD Performance Budgeting Survey, Question 30, OECD, Paris

Steering and coordinating policy evaluation - the role of the centre of government

The CoG is known by different names in different countries, such as the Chancellery, Cabinet Office, Office of the President, Office of the Government, etc. It plays an increasingly active role in policy development, co-ordination and monitoring across public administration. The CoG aims to secure a strong, coherent and collective strategic vision - especially as it relates to major cross-departmental policy initiatives (OECD, 2014^[92]). Its role can be crucial in policy evaluation across government, as it requires co-ordination across different departments and ministries.

According to the OECD survey on the Organisation and Functions of the Centre of Government (2017), policy co-ordination across government and monitoring the implementation of government policy is among the five key responsibilities of the CoG across the OECD.

The OECD survey on policy evaluation also finds that the CoG plays a crucial role in embedding a whole-of-government approach to policy evaluation. An assessment of the mandate of countries' CoG reflects its role as a guiding institution in policy evaluation across government (Table 2.3). In 16 OECD countries (18 countries in total), the CoG's mandate includes the definition and update of the evaluation, while in 15 countries (including 14 OECD countries) it includes providing incentives for carrying out policy evaluation. In 19 of the 23 OECD countries (21 of the 27 countries in total) in which the CoG has a role in policy evaluation across government, it is tasked to promote the use of policy evaluation. The CoG is also in charge of providing guidelines for policy evaluation. Only seven OECD countries (10 in total) mentioned that the CoG is responsible for defining the course of action for commissioning policy evaluation.

Table 2.3. Mandate of Centre of Government for policy evaluation

Note: n=27 (23 OECD member countries). Answers reflect responses to the question, "Please list the duties and responsibilities of this/these

	Defining and updating the evaluation	Developing guideline(s)	Providing incentives for carrying out policy evaluations	Undertaking policy evaluations	Requiring government institutions to undertake specific policy evaluations	Defining course of action for commissioning evaluations	Developing skills, competences and/or qualifications of evaluators	Developing standards for ethical conduct	Ensuring quality standards of evaluations	Promoting stakeholder engagement in evaluations	Overseeing the evaluation calendar and reporting	Promoting the use of evaluation	Serving as a knowledge centre and providing a platform for exchange	Following up on evaluation reports
Australia	○	○	○	●	●	●	○	○	○	○	○	●	○	●
Canada	●	●	○	●	●	○	●	●	○	○	●	●	●	●
Estonia	●	●	●	●	○	○	●	○	○	●	○	●	○	●
Finland	○	●	●	○	○	●	○	●	●	○	●	●	●	●
France	●	●	●	●	●	○	●	○	●	●	●	●	○	○
Germany	●	●	●	●	○	○	●	○	●	●	●	●	●	●
Great Britain	.	f.	.	.	.	.	.	.	.	.	.	.	.	.
Greece	●	●	○	○	●	●	●	○	●	●	●	●	○	●
Hungary	●	●	●	●	●	○	○	○	○	○	○	●	●	●
Iceland	○	○	○	○	○	○	○	○	○	○	○	○	●	○
Israel	●	●	●	○	●	●	●	○	●	●	●	●	●	○
Italy	●	●	●	●	●	●	○	●	○	●	○	●	●	○
Korea	●	●	●	●	●	○	●	●	●	●	●	●	●	●
Latvia	●	●	●	●	●	○	●	○	●	●	○	●	●	●
Lithuania	●	●	●	○	●	○	●	○	○	●	○	●	●	○
Netherlands	○	○	○	○	○	○	○	○	○	○	○	●	○	○
New Zealand	○	○	○	○	○	○	○	○	○	○	○	●	○	○
Portugal	●	●	●	●	●	●	●	●	●	●	●	●	●	●
Slovakia	○	○	●	●	○	○	●	○	○	○	○	●	●	○
Slovenia	●	○	●	○	●	●	○	○	●	○	●	●	○	●
Spain	●	○	○	○	○	○	○	○	○	○	○	○	○	○
Turkey	●	○	○	●	●	○	○	○	○	○	○	○	○	○
United States	●	●	●	○	●	○	●	○	●	●	○	●	●	○
OECD Total														
● Yes	16	14	14	12	14	7	12	5	10	11	9	19	13	11
○ No	6	8	8	10	8	15	10	17	12	11	13	3	9	11
Argentina	●	●	○	●	●	●	○	○	●	●	●	●	●	●
Brazil	○	●	●	○	●	●	○	○	○	●	○	○	○	○
Costa Rica	●	●	○	○	●	●	○	○	○	○	○	●	○	●
Romania	○	●	○	○	○	○	○	○	●	○	●	○	○	○

institution/s related to policy evaluation across government" for the Centre of Government / Presidency / Prime Minister's Office / Cabinet Office or equivalent. Answer option "Other" is not included. Note: The UK approach to policy evaluation splits these responsibilities amongst the Cabinet Office, the Treasury, and professional analysts across government (e.g. the Government Economic Service and the Government Social Research Service), with most institutions also developing their own supplementary guidance and some form of ministerial/management response to the results. Source: OECD (2018) Survey on Policy Evaluation

In sum, the CoG has a vital role to play in providing strategic direction for policy evaluation as well as incentives for other institutions to use the evaluation findings (See Chapter 3.). CoG institutions can facilitate policy evaluation across government, due to the centre's role in steering and coordination. This is consistent with the OECD cross-country analysis of CoG functions, which concludes that “more collaborative strategies for achieving policy goals suggest a role for the centre that is less about being a watchdog or internal auditor and more about providing active facilitation, support and implementation advice to ministries or groups of ministries. This is especially the case for meeting cross-cutting policy goals” (*Centre Stage 2 report - The organisation and functions of the centre of government in OECD countries* (OECD, 2018^[93])).

The close proximity to strategic decision making yields a number of benefits for the role of CoG in relation to promoting evaluation. Allocating the role of principal institution in charge of policy evaluation close to political power can be interpreted as a sign of political commitment. In Germany for example, the main institution in charge of policy evaluation is located in the Chancellery (*Bundeskanzleramt*) which has government-wide co-ordination powers. In addition, the CoG usually has the political leverage to ensure that the findings of evaluations are subsequently used in forthcoming steps of the policy-cycle. Still, this may require to balance the trade-offs between influence and the perception of independence of the evaluation.

An assessment of the role of the CoG also requires discussing who takes responsibility for policy evaluation. In a majority of countries (17 countries, 16 OECD countries), civil servants are head of the policy evaluation unit within CoG. Surprisingly, only in nine countries (of which six are OECD countries) such as Greece and Hungary, political appointees are given such a role. In all these nine countries apart from Israel, employees appointed for heading evaluation are replaced when government changes. On the contrary, civil servants responsible for evaluation in CoG are rarely replaced.

The CoG majorly finances its evaluation units through its own budget with 19 countries, including 15 OECD countries, doing so. On the other hand, only five countries, such as Finland, directly and independently allocate funds from the national budget to their CoG's evaluation units. Argentina is the only country where this unit (National Directorate for Information System, Monitoring and Evaluation of Social Programs, SIEMPRO) is financed through both the CoG's budget and the national budget. In contrast, the Slovakian evaluation unit is financed by European structural and investment funds.

Ministry of Finance / Ministry of Economy / Ministry of Treasure or equivalent

In many countries, the institutionalisation of policy evaluation originated in economic incentives with the aim to enhance the quality of public expenditures and improve results in terms of government's spending. In Australia for example, the ministry of finance is responsible for the spending review procedures. In the OECD survey, it was mentioned as second most frequent institution (26 countries in total, including 22 OECD countries) with competences for policy evaluation across government (see Box 2.11 for details on the Netherlands).

Box 2.11. The role of the Ministry of Finance in institutionalising policy evaluation: the experience of the Netherlands'

The Dutch government launched an “Insight into Quality” Plan following the coalition agreement for 2017-21. This initiative aims to increase the knowledge concerning policies’ efficiency and effectiveness. Accordingly, the ministry of finance works in coordination with all departments to understand and strengthen the government-wide structure of the country’s evaluation system and the added-value of its policies for citizens. To achieve these goals, the following initiatives have been established:

- The ministry of finance started monitoring and ensuring the application of the article associated to the revised Budget Law of January 2018 to strengthen the evaluation system. It will work on a proposal to this effect with other ministries, explicitly involving the parliament as recipient of the information in this process.
- The ministry of finance also stimulates mutual learning across ministries through interdepartmental seminars.
- Following the initiatives of other ministries, the ministry of finance has set up a policy quality and evaluation committee, composed of core department officials, agencies and external experts to increase the internal attention to and quality of policy evaluations.
- The finance ministry worked with other ministries to update the **Integrated Assessment Framework** by focussing it on effectiveness and efficiency.

The ministry ensures **compliance to rules on evaluation of fiscal policy**, guaranteeing that these are carried out and coordinating interdepartmental policy studies.

Ministries of finance have been identified as the principal actors to undertake policy evaluations, in 15 OECD countries and 18 countries in total. They less frequently have the role to define the course of action for commissioning evaluations, which was mentioned by only two countries. Ministries of finance have developed (or are developing) guidelines for policy evaluations across government in 13 countries including 12 OECD countries. In 11 countries, including eight OECD countries, the ministries of finance are tasked to follow up on evaluation reports. These findings reflect the ministries’ role as coordinating institution that provides guidance and in some cases supervision to other ministries’ evaluation activities.

Box 2.12. Key practices for value-for-money assessment- a key responsibility for the Ministry of Finance or equivalent

The 2019 OECD Report Budgeting and Public Expenditures in OECD Countries assessed key practices for value-for-money assessments. The Report finds the following:

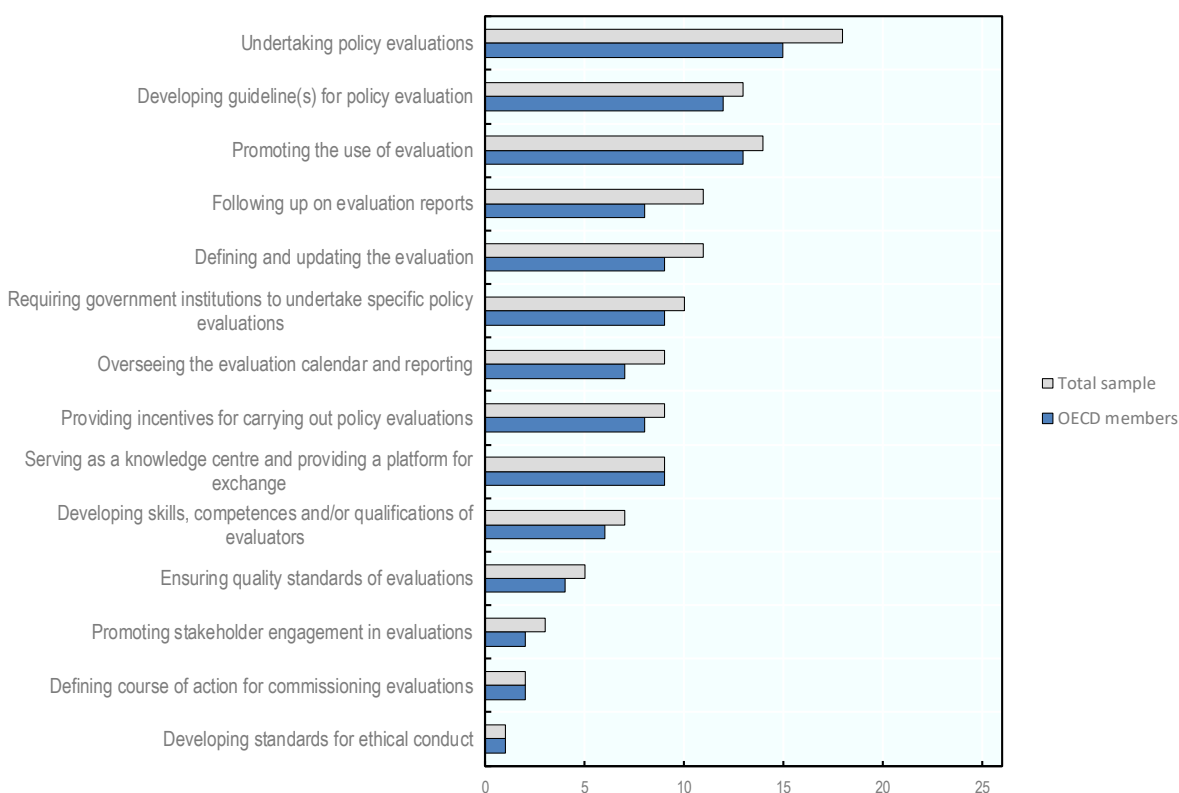
"In principle, the general assessment of costs and benefits of an investment should be the driving force for the prudent evaluation of investment decisions (OECD, 2015). Value for money (VfM) can be defined as what a government judges to be an optimal combination of quantity, quality, features and price (i.e. cost), expected over the whole of the project's lifetime. VfM can be measured in absolute cost-benefit terms (Do the benefits exceed the costs?) or in relative terms (Is one form of delivery more cost-effective than the other – see next section). In many cases, VfM is assessed using a combination of quantitative (such as cost/benefit analysis) and qualitative tools. A majority of surveyed countries conduct both absolute and relative VfM assessments for either all of the projects or for those projects above a certain threshold, no matter whether the projects are delivered via PPPs or traditionally procured. In some, countries, such as Slovakia and Austria VfM assessments are only compulsory for some line ministries (e.g. railways in Austria).

There are several techniques for assessing value for money. Cost-benefit analysis (including total cost of ownership during the life-cycle) is the most popular approach (89%), followed by net present value (70%) and cash-flow estimates over the project cycle (70%). About half of the countries also use other tools, including internal rate of return, analysis of the willingness of users to pay, or business case methodology. In many cases these VfM assessments are assessed by combined approaches. Denmark for example, calculates and reports the socio-economic value and conducts business cases. Norway follows an alternative approach assessing all PPP projects and all large investment projects (over NOK 750 million) within a general quality assurance scheme."

Source: OECD (2019^[12]), Budgeting and Public Expenditures in OECD Countries 2019, <https://dx.doi.org/10.1787/9789264307957-en>.

Similarly to the CoG, the ministries of Finance's mandate includes the promotion of the use of evaluation findings. This finding can be embedded in processes led by ministries of finance, such as ensuring value-for-money or feeding into budget related processes such as spending reviews (see Figure 2.5).

Figure 2.5. Mandate of the Ministry of Finance/Economy/Treasury or equivalent



Note: n=26 (22 OECD member countries). Answers reflect responses to the question, "Please list the duties and responsibilities of this/these institution/s related to policy evaluation across government" for the Ministry of Finance / Ministry of Economy / Ministry of Treasury or equivalent. Answer option "Other" is not included.

Source: OECD Survey on Policy Evaluation (2018)

Of the 26 countries, which noted that the ministry of finance or equivalent has a role in policy evaluation across government, 19 countries also mentioned the CoG as well. Still this may reflect responsibilities with different parts of the evaluation systems, for example with centres of government leading on evidence informed policy making processes and the coordination of the regulatory process, while Ministries of finance are more concerned with budgeting and expenditure management. In any case, there is a need for a clear-cut allocation of responsibilities in order to ensure alignment.

Further analysis shows that evaluation units within ministries of finance are mostly headed by civil servants. This is the case in 19 out of the 26 countries in which the Ministry of Finance has a mandate relating to policy evaluation, including 16 OECD countries. Similarly to the patterns observed in CoG, civil servants responsible for evaluation are usually replaced when the government changes in Spain and Brazil. Some Ministries of Finance appoint political employees as heads of their evaluation units, as is the case in six countries such as Romania. Half of such political appointees are replaced when the government changes (Chile, Hungary, and Mexico).

More than half of evaluation units hosted by ministries of finance are financed by the budget of the ministry itself, as is the case in 16 countries (13 in the OECD). Only five countries, including four OECD countries (Germany, Finland, the Netherlands and Norway) allocate national budget to the evaluation units within their ministries of finance. In contrast, Lithuania's Economic Analysis and Evaluation Unit, hosted in the ministry of finance and responsible for the evaluation of EU structural funds, is financed through Technical Assistance allocations.

Ministry of Planning, Development, or equivalent

Not all of the 42 countries surveyed for this Report have established a dedicated Ministry for Planning, Development or equivalent that has competences related to policy evaluation across government (7 countries in total, including 4 OECD countries: Chile, Czech Republic, Poland and Slovakia). In Latin American countries such as Brazil, Colombia and Costa Rica (see Box 2.13 for details on Brazil and Colombia), the ministries of Planning takes an active role in policy evaluation. Latin American countries make up four of the countries in which a ministry of planning, development or equivalent has an active role in policy evaluation across government which can be traced back to the strong role that national development plans have in the region. These strategic plans tend to be evaluated by the ministries of planning or development, which gives them a mandate for policy evaluation across government.

Box 2.13. Committee for Monitoring and Evaluation of Federal Public Policies (CMAP) in Brazil & Colombia's National Planning Department

The **Committee for Monitoring and Evaluation of Federal Public Policies (CMAP)** was created in 2016 under the co-ordination of the Brazilian ministry of planning. It has the objective to improve the actions, programmes and public policies of the federal executive branch, as well as the allocation of resources and the quality of public spending. The CMAP regroups representatives of the ministries of planning, budget and management, ministry of finance, the Civil House of the Presidency of the Republic and the office of the Comptroller general of the Union, with special participation of members of public and private institutions.

Its role is to define the policies, programmes and actions that will be monitored and evaluated, and propose guidelines to improve them by using thematic committees. Moreover, the committee makes recommendations to policy makers on the adoption, adjustments and improvements of policies, under principles of transparency and accountability.

Colombia's National Planning Department (DNP) is explicitly entrusted national planning responsibilities. The department encompasses the National Public Management Results Evaluation System (SINERGIA) and the Public Policy Monitoring and Evaluation Division (DEPP). Through SINERGIA and the DEPP, DNP has gradually solidified the incorporation of evaluation in government-wide policy implementation. SINERGIA also represents the role international organisations can play in institutionalisation, as the Inter-American Development Bank provided assistance in the development of SINERGIA (Lazaro, 2015^[5]).

Source: Diário oficial da Uniao (2016), "Portaria interministerial nº 102", 7 April 2016.

Figure 2.6. Mandate of the Ministry of Planning, Development, or equivalent



Note: n=7 (4 OECD member countries). Answers reflect responses to the question, "Please list the duties and responsibilities of this/these institution/s related to policy evaluation across government" for the Ministry of Planning, Development, or equivalent. Answer option "Other" is not included.

Source: OECD Survey on Policy Evaluation (2018).

In four countries, including Chile, Brazil, Colombia and Costa Rica, the ministry of planning, development or equivalent, finances its evaluation activities from its own budget. In the three other countries where this ministry has evaluation-related responsibilities, evaluation units are financed through other means, such as EU funds as in Poland.

The role of autonomous agencies

In addition, autonomous agencies have taken up competences related to policy evaluation across government in some countries. One example of such autonomous agency with a role in evaluation is Mexico's National Council for the Evaluation of Social Development Policy, CONEVAL (see Box 2.14). Another example of a country with several autonomous agencies contributing to policy evaluation is Italy (Box 2.15).

In Denmark, autonomous agencies can have an *ad hoc* role across government when the evaluation of a specific policy is requested by the parliament. However, there is no institution within the executive who by default have competences related to policy evaluation across the Danish government.

Box 2.14. The National Council for the Evaluation of Social Development Policy of Mexico (CONEVAL)

CONEVAL's mandate encompasses between 100 to 130 federal programmes from year to year, all of which are required to execute internal evaluations governed by CONEVAL's guidelines. The organisation also directly oversees over a dozen evaluations per year. Results from the evaluations are influential. In 2013-2014, half of evaluated programs were substantially refocused and 41 percent of programs underwent corrections of activities or operational aspects (Lázaro, 2015^[54]).

CONEVAL is the main vehicle towards the institutionalisation of policy evaluation within Mexico through initiatives that clarify, cement, and advance M&E processes. The 2007 issuance of the mandatory General Guidelines for federal programme evaluations provided definitions, regulations, principles, and requirements for components of the monitoring and evaluation system. In 2008, a tracking system for evaluations was implemented and subsequent efforts were made to develop that system onto digital platforms and make it accessible to the wider public. Training seminars were also organised for programme managers (Gaarder and Briceño, 2010^[55]). These efforts may have the long-term effects of structuring evaluation practices and increasing capacity, even outside of CONEVAL affiliated entities—embedding an evaluation culture (Lázaro, 2015^[54]).

Sources: (Lázaro, 2015^[54]), (Gaarder and Briceño, 2010^[55]).

Box 2.15. The role of autonomous agencies in Italy

Italy offers a good examples of the wide distribution of roles for policy evaluation, which can be given to autonomous agencies, acting either as knowledge brokers, or as part of their duty as regulators, to assess regulatory impacts. Overall the functions remain sectoral.

A number of autonomous agencies have their own legislative framework requiring them to perform policy evaluation, which *de facto* gives them a knowledge brokerage role. There are no less than 3 agencies in the education area, including INDIRE, the National Institute for Documentation, Innovation and Educational Research, which is the oldest research organisation related to the Italian Ministry of Education, INVALSI, the national institute for the evaluation of the education and training system and ANVUR for higher education, the national institute for evaluation of universities and research. In the area of labour and social inclusion, INAPP, the national institute for public policy analysis.

In addition, some other agencies have sectoral responsibilities, which include tasks in terms of monitoring: an agency like ANPAL, for active labour market policies has responsibility for the analysis monitoring and evaluation of active labour market policies, more in terms of quantifying indicators on the degree of achievement of the annual objectives for ALPs and monitoring the expected results.

In addition, regulators, such as the CONSOB, the Securities and Exchange Commission in Italy (CONSOB) are also practicing evaluations when introducing new regulations. The example of the regulation on equity crowdfunding was provided to the OECD when developing the Policy Framework (Impact Assessment Office, 2018^[94]). Crowdfunding provides alternatives to bank loans as the supply of bank loans dwindled during the financial crisis. The evaluation included a mapping of burdens and a qualitative analysis of the costs benefits analysis. The evaluation contributes to the regulatory impact assessment and allows to check whether the objectives intended by legislators have been achieved.

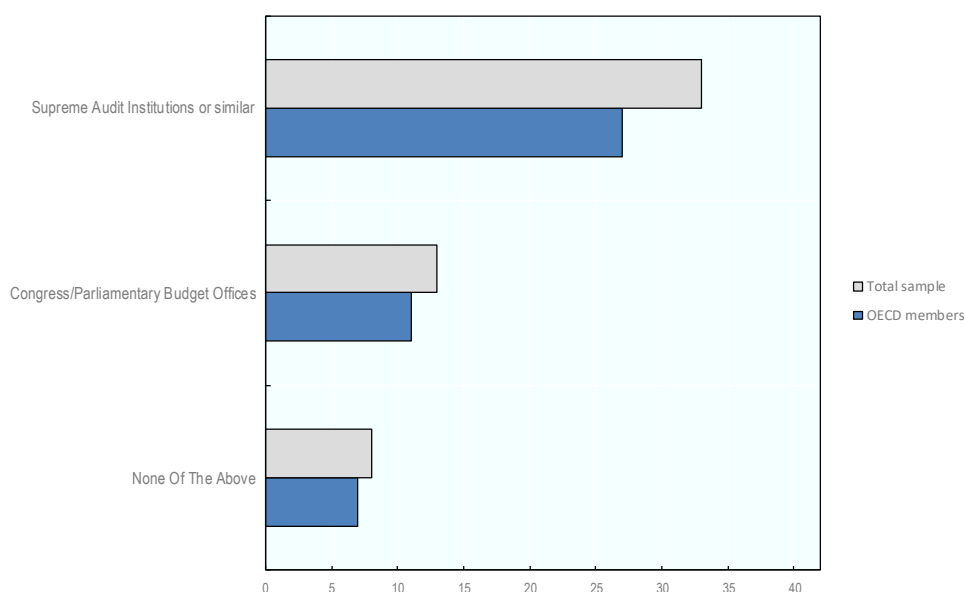
Source: Italian submissions to the Secretariat.

Taking up the discussion on the trade-off between political independence and influence, autonomous agencies tend to be more independent than other institutions or ministries. The prevalence of the agencies' self-determined evaluation agendas contributes to their independence. In Mexico, the United States and Costa Rica, where autonomous agencies have a role in policy evaluation across government, they are financed through the budget allocation of the hosting institution.

Institutions beyond the Executive

A number of actors and institutions outside the executive have a crucial role in policy evaluation and its institutionalisation. In OECD countries, 27 Supreme Audit Institutions (SAIs) (33 in all countries) have competences on policy evaluation at central/federal level (Figure 2.7). Around one third (11) of OECD countries (13 countries in total) involve Congress or the Parliamentary Budget Office in policy evaluation. In seven OECD countries (8 countries in total), none of the aforementioned institutions has a mandate on policy evaluation. In addition to performing specific evaluations, SAI can also play a useful role to offer general guidance in evaluating the evaluation system as a whole.³

Figure 2.7. Institutions beyond the executive that have competences for policy evaluation at central/federal level



Note: n=42 (35 OECD member countries). Answers reflect responses to the question, "Which of the following institutions beyond the executive have competences on policy evaluation at central/federal level? (Check all that apply)".

Source: OECD Survey on Policy Evaluation (2018).

The OECD is closely collaborating with SAIs and has assessed their roles in detail. Among others, the OECD published the Good Practices in Supporting Supreme Audit Institutions in 2010 (OECD, 2010_[95]). In 2016, the OECD launched a Report on *Supreme Audit Institutions and Good Governance- Oversight, Insight and Foresight* (OECD, 2016_[21]). The report mapped the activities of ten leading Supreme Audit Institutions (SAIs) in Brazil, Canada, Chile, France, Korea, the Netherlands, Poland, Portugal, South Africa and the United States (see Box 2.16 for details on Chile). In particular, it examined how these SAIs assess key stages of the policy cycle, and provided examples and case studies of SAIs' activities, supporting the integration of international good practices into policy formulation, implementation and evaluation.

Box 2.16. Chile's Supreme Audit Institution's role in strengthening good governance

In 2014, the OECD conducted a Public Governance Review of the SAI of Chile. The Report finds that “Chile's supreme audit institution (*Contraloría General de la República de Chile* or CGR) is at the forefront of an evolution of Supreme Audit Institutions and has undertaken ambitious initiatives for institutional strengthening, capacity development, transparency and citizen participation. The CGR has introduced strategic planning, restructured its workforce and become an exemplary institution with respect to transparency within the Chilean public sector”.

The CGR has a role to enhance good public governance, and improve accountability and the quality of government decision-making. The CGR can provide objective and credible information that is widely recognised as useful.

Source: (OECD, 2014^[96]).

In addition, Supreme Audit Institutions can play a significant policy evaluation function in a number of countries, even if these activities may crossover with performance audits in some cases. Parliament can request evaluations, or performance audits with an evaluative approach. Good practices exist in several countries, including Switzerland, the United States and France. In Switzerland, the Federal Audit Office has developed a specific competence centre for Evaluation,⁴ which includes professional evaluators and follows international guidelines of the supreme audit institutions in the area (ISSAI 300) as well as the standards of the Swiss Evaluation Society. In the United States, the U.S. Government Accountability Office (GAO) is an independent, nonpartisan agency that works for Congress, which provides Congress and federal agencies with objective, reliable information to help the government save money and work more efficiently. The GAO demonstrates many best practices in evaluation, in particular the accessibility of its reports, and the clarity of its line of inquiry, together with a wealth of analytical results. Finally, in France, the Supreme Audit Institution has also received an official mandate for evaluation through the Constitution (Box 2.17).

Box 2.17. The French Supreme Audit Institution's role in conducting policy evaluation

In 2008, the role of the French Supreme Audit Institution (Cour des Comptes) in evaluating public policies was embedded in the French Constitution. In 2011, this constitutional competence was translated into law, enabling the French SAI to conduct evaluations either at the request of Parliament or of its own accord.

The Cour des Comptes operates in accordance with INTOSAI guidelines adopted in 2016 (mentioned earlier) as well as with the specific professional standards adopted in 2014. To date, it has carried out and published over 20 evaluations on an array of subjects ranging from health to housing to education.

In conducting its evaluations, the Cour des Comptes leverages both quantitative and qualitative data and collaborates with external laboratories to carry out the analysis. It also actively involves relevant stakeholders in the evaluation exercises.

Source: Input from the Cours des Comptes (France)

The INTOSAI's working group on programme and public policy evaluation conducted a survey on the implementation of INTOSAI GOV 9400 guidelines (See Box 2.18). The results show that 31% of responding SAIs perform public policy and programme evaluations. Over 60% of SAIs indicated that

they only conduct performance audits, without an evaluative approach, whereas 6% carried out performance audits with an evaluative approach. Overall discussions among the SAIs community reflect this duality of roles, between audit strictly speaking and evaluation and the different professional cultures, skills and approaches that the two may require.

Box 2.18. A cross-country perspective at Supreme Audit Institutions' role in policy evaluation

The International Organisation of Supreme Audit Institution's (INTOSAI) has issued the INTOSAI GOV 9400 Guidelines on evaluation of public policies and monitoring its implementation through its working group on evaluation. These Guidelines seek to harness this potential by providing quality standards to enable SAIs to appropriately select topics to be evaluated, involve stakeholders and experts, plan evaluations, choose tools and methods, and apply and publish reports. These guidelines reiterate that given their independent institutional position, grasp of evaluation methodologies, and knowledge of public policies, SAIs are naturally suited for evaluating public policies.

Note: The guidelines are: www.intosaicommunity.net/wgeppp/wp-content/uploads/2019/08/INTOSAI-GOV-9400_ENG.pdf
Source: (INTOSAI and Cour des Comptes (France), 2019^[97]), Input from Cour des Comptes (France)

Moreover, although in a majority of respondent countries, the SAIs – alongside parliament – can launch policy evaluations, only 40% of respondents indicated that they conducted more than three evaluations per year. These findings can be attributed to resource and time constraints. The most common difficulties that are reported for SAIs in conducting evaluations include: the timeframe for carrying out evaluations (22%), the use of methodological tools (16%), and insufficient quantitative human resources (13%). Institutional factors, such as the degree of understanding and acceptance of the evaluation process within the SAI, may also create obstacles. Parliamentary Budget Offices can also have a role (see (OECD, 2019^[12])).

The rapid growth of independent fiscal institutions, including independent parliamentary budget offices and fiscal councils, gives them an important role, as one of their key function is to “produce, assess and/or endorse macroeconomic or fiscal forecasting, monitoring compliance with fiscal rules, policy costing, long-term fiscal sustainability analysis, and supporting the legislature in budget analysis” (OECD, 2019^[12]).

In sum, different actors beyond the executive branch of government have significant roles in policy evaluation. Evaluations that are internal or external to the executive branch carry different functions and may yield different contributions. External evaluations provide greater provisions for transparency and accountability but offer less scope for promoting use as an internal management tool from the government's centre, as when this is coordinated for example through the centre of government, a ministry of finance or equivalent (budget central authority, planning, presidency or internal control office). The two types of evaluations should probably be seen as complementary to one another.

Non institutional actors and International Organisations

The report focuses on the institutional actors and mechanisms that may exist as part of the public sector capacity to promote use and quality of evaluation. However, two types of actors were not addressed through the survey and may play a significant role in countries.

The first includes non-governmental organisations (NGOs) and citizens. NGOs often play a significant role as suppliers of evaluation, to drive and push the issues that are part of their core mission. Equally, they are also often strong users of evaluation products that can serve their purpose. In some countries, particularly those with a less developed public sector, NGOs can help fill the gap of the evaluation

ecosystem, ensuring some data collection and providing some evaluations that help to inform policy decisions. In addition, citizens have a significant role to play, to not only inform and engage in evaluation processes, but also as a group to promote the use of evaluations and decisions that can make a difference in their daily lives. There is a full topic of citizen engagement in the evaluation processes and the role of public deliberation that may go beyond the scope of the current report, but which is worth mentioning. They are certainly a significant part of a healthy evaluation ecosystem.

The second type of actors is the International Organisations (IOs). IOs, including the OECD, play a significant role in promoting evaluation and peer learning at the domestic level. Various international organisations may have different mandates and functions, some more in the economic and financial domain, and some more geared towards specific topics. OECD work as such also often has an evaluative nature, as it takes a cross-country approach to assess policy outcomes and identify best practices. The question is the extent to which international organisations could have an impact on the evaluation ecosystem as a whole at country level. For example, the OECD recently completed a full review of the Irish Government Economic Evaluation Service and is conducting country specific work in a few other countries. (OECD, 2020^[98])

Coordination mechanisms

Coordination bodies or mechanisms such as commissions and integrated services enable aligning and sharing practices across institutions within and beyond government, which is a necessary disposition for fostering a sound evaluation culture. For example, Mexico's National Council for the Evaluation of Social Development Policies (CONEVAL) aims to improve coordination for better evaluation activities across government and across states. In the United States, an interagency council has been set up that regroups Evaluation Officers and is intended to serve as a forum for exchanging information and advising the Office of Management and Budget on issues affecting the evaluation functions such as the evaluator competencies, best practices for programme evaluation, and evaluation capacity building.

Another example of an integrated cross-government service for building analytic capacity for evaluation to improve policymaking is the Irish Government Economic and Evaluation Service (IGEES). The IGEES is a horizontal structure coordinated by the Department of Expenditure and Reform that offers support to the whole Irish Government in delivering evidence-informed policy making. IGEES staff are working across all departments. (OECD, 2020^[98])

However, 18 countries (including 15 OECD countries) do not have regular consultation on policy evaluation issues between the government (executive) and SAIs. Although 12 countries (including 10 OECD countries) do not have consultation mechanisms, they conduct regular *ad hoc* consultation, as is the case for example in Austria, Czech Republic, Denmark, Greece, Finland, and Ireland. Nevertheless, only seven countries (including 5 OECD countries) confirmed the existence of a formal co-ordination mechanism for regular consultation, such as Estonia and Hungary. Lithuania, Latvia, and Estonia have a formal co-ordination mechanism to avoid overlaps on planned or on-going evaluations.

Notes

¹ Centre of government is defined as an administrative structure that serves the Executive (President or Prime Minister, and the Cabinet collectively). For further information about CoG, see sub-section *Institutions within the Executive*.

² Inspektionen för socialförsäkringen (The Swedish Social Insurance Inspectorate), Kulturanalys (The Swedish Agency for Cultural Policy Analysis), Tillväxtanalys (The Swedish Agency for Growth Policy Analysis), Trafikanalys (Transport Analysis), Vårdanalys (The Swedish Agency for Health and Care Services Analysis), Brå (The Swedish National Council for Crime Prevention) and IFAU (Institute for evaluation of labour market and education policy).

³ See Federal Auditor report on the capacity of federal services to evaluate public services www.ccrek.be/Docs/2018_09_CapaciteServicesPublicsFederauxAEvaluerLesPolitiquesPubliques.pdf

⁴ <https://www.efk.admin.ch/fr/ueber-uns/organisation/centres-de-competence/1262-fb6-f.html>

References

- 115th Congress (2019), *Public Law No: 115-435 (01/14/2019) - Foundations for Evidence-Based Policymaking Act of 2018*, <https://www.congress.gov/bill/115th-congress/house-bill/4174>. [7]
- Acquah, D., K. Lisek and S. Jacobzone (2019), “The Role of Evidence Informed Policy Making in Delivering on Performance: Social Investment in New Zealand”, *OECD Journal on Budgeting*, Vol. 19/1, <https://dx.doi.org/10.1787/74fa8447-en>. [18]
- AEVAL (2015), *Practical guide for the design and implementation of public policy evaluations* (Guía práctica para el diseño y la realización de evaluaciones de políticas públicas), http://www.aeval.es/export/sites/aeval/comun/pdf/evaluaciones/Guia_Evaluaciones_AEVAL.pdf (accessed on 21 August 2019). [131]
- Alkin, M. and S. Taut (2002), “Unbundling evaluation use”, *Studies in Educational Evaluation*, Vol. 29/1, pp. 1-12, [http://dx.doi.org/10.1016/S0191-491X\(03\)90001-0](http://dx.doi.org/10.1016/S0191-491X(03)90001-0). [161]
- American Evaluation Association (2018), *Guiding Principles*. [148]
- American Evaluation Association (2015), *Core Evaluator Competencies*, <http://www.eval.org>. [151]
- Barnett, C. and L. Camfield (2016), “Ethics in evaluation”, *Journal of Development Effectiveness*. [117]
- Better evaluation (2019), *Review evaluation (do meta-evaluation)*, https://www.betterevaluation.org/en/rainbow_framework/manage/review_evaluation_do_meta_evaluation (accessed on 19 August 2019). [142]

- Bossuyt, J., L. Shaxson and A. Datta (2014), "Study on the uptake of learning from EuropeAid's strategic evaluations into development policy and practice", Evaluation Unit of the Directorate General for Development and Cooperation-EuropeAid (European Commission). [67]
- Bridgeland, J. and P. Orszag (2013), *Can Government Play Moneyball? - The Atlantic*, <https://www.theatlantic.com/magazine/archive/2013/07/can-government-play-moneyball/309389/> (accessed on 6 December 2018). [165]
- Brown, L. and S. Osborne (2013), "Risk and Innovation", *Public Management Review*, Vol. 15/2, pp. 186-208, <http://dx.doi.org/10.1080/14719037.2012.707681>. [60]
- Brown, R. and D. Newman (1992), "Ethical Principles and Evaluation Standards: Do They Match?", *Evaluation Review*, Vol. 16/6, pp. 650-663, <http://dx.doi.org/10.1177/0193841X9201600605>. [119]
- Bundesministerium für Finanzen and Bundesministerin für Frauen und öffentlichen Dienst (2013), *Handbuch Wirkungsorientierte Folgenabschätzung Arbeitsunterlage*, <http://www.oeffentlicherdienst.gv.at> (accessed on 28 August 2019). [152]
- Campbell, S. and G. Harper (2012), *Quality in policy impact evaluation: understanding the effects of policy from other influences (supplementary Magenta Book guidance)*, HM Treasury, <http://www.nationalarchives.gov.uk/doc/open-> (accessed on 9 July 2019). [129]
- Canada Treasury Board (2016), *Policy on results*, <https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=31300> (accessed on 20 September 2019). [8]
- Caroline Heider (2018), *The Three Pillars of a Working Evaluation Function: IEG's Experience*, <http://ieg.worldbankgroup.org/blog/three-pillars-working-evaluation-function-iegs-experience> (accessed on 22 August 2019). [133]
- Cinar, E., P. Trott and C. Simms (2018), "Public Management Review A systematic review of barriers to public sector innovation process", <http://dx.doi.org/10.1080/14719037.2018.1473477>. [65]
- Commission on Evidence-Based Policymaking (2017), *The Promise of Evidence-Based Policymaking: Report of the Commission on Evidence-Based Policymaking*, <https://www.cep.gov/report/cep-final-report.pdf> (accessed on 6 August 2019). [31]
- CONEVAL (2007), *Lineamientos generales para la evaluación de los Programas Federales de la Administración Pública Federal*, https://www.coneval.org.mx/rw/resource/coneval/eval_mon/361.pdf (accessed on 18 June 2019). [23]
- Cooksy, L. and M. Mark (2012), "Influences on evaluation quality", *American Journal of Evaluation*, Vol. 33/1, pp. 79-84, <http://dx.doi.org/10.1177/1098214011426470>. [159]
- Crowley, D. et al. (2018), "Standards of Evidence for Conducting and Reporting Economic Evaluations in Prevention Science", *Prevention Science*, Vol. 19/3, pp. 366-390, <http://dx.doi.org/10.1007/s11121-017-0858-1>. [45]
- Damschroder, L. et al. (2009), "Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science", *Implementation Science*, Vol. 4/1, p. 50, <http://dx.doi.org/10.1186/1748-5908-4-50>. [109]

- Departamento Nacional de Planeación (2016), *¿Qué es una Evaluación?*, [25]
<https://sinergia.dnp.gov.co/Paginas/Internas/Evaluaciones/%C2%BFQu%C3%A9-es-Evaluaciones.aspx>.
- Department of the Prime Minister and Cabinet (2014), *Guide to Implementation Planning*, [42]
 Australian government, <https://www.pmc.gov.au/sites/default/files/files/guide-to-implementation-planning.pdf> (accessed on 12 July 2019).
- Department of the Prime Minister and Cabinet (Australia) (2013), *Policy Implementation* |, [85]
<https://www.pmc.gov.au/government/policy-implementation> (accessed on 20 September 2019).
- Desautels, G. and S. Jacob (2012), “The ethical sensitivity of evaluators: A qualitative study using a vignette design”, *Evaluation*, Vol. 18/4, pp. 437-450, [116]
<http://dx.doi.org/10.1177/1356389012461192>.
- Dobbins, M. et al. (2009), “A randomized controlled trial evaluating the impact of knowledge translation and exchange strategies”, *Implementation Science*, Vol. 4/1, p. 61, [176]
<http://dx.doi.org/10.1186/1748-5908-4-61>.
- Estonian National Audit Office (2011), *The state of affairs with the legislative impact assessment*. [157]
- European Commission (2017), *Better Regulation Guidelines*, http://europa.eu/about-eu/basic-information/decision-making/treaties/index_en.htm. [171]
- European Court of Auditors (2013), *Audit Guidelines on Evaluation*, [158]
https://www.eca.europa.eu/Lists/ECADocuments/GUIDELINES_EVALUATION/Evaluation-Guideline-EN-Oct2013.pdf (accessed on 23 August 2019).
- European Environment Agency (2017), “EEA guidance document-policy evaluation”, [27]
<https://www.researchgate.net/publication/317594615>.
- Flay, B. et al. (2005), “Standards of Evidence: Criteria for Efficacy, Effectiveness and Dissemination”, *Prevention Science*, Vol. 6/3, pp. 151-175, [50]
<http://dx.doi.org/10.1007/s11121-005-5553-y>.
- Fleischer, D. and C. Christie (2009), “Evaluation use: Results from a survey of U.S. American evaluation Association members”, *American Journal of Evaluation*, Vol. 30/2, pp. 158-175, [162]
<http://dx.doi.org/10.1177/1098214008331009>.
- Flemig, S., S. Osborne and T. Kinder (2016), “Risky business—reconceptualizing risk and innovation in public services”, *Public Money & Management*, Vol. 36/6, pp. 425-432, [61]
<http://dx.doi.org/10.1080/09540962.2016.1206751>.
- France Stratégie (2016), *How to evaluate the impact of public policies: a guide for the use of decision makers and practitioners (Comment évaluer l’impact des politiques publiques : un guide à l’usage des décideurs et des praticiens)*, [128]
https://www.strategie.gouv.fr/sites/strategie.gouv.fr/files/atoms/files/guide_methodologique_20160906web.pdf (accessed on 21 August 2019).
- France Stratégie, R. Desplat and M. Ferracci (2016), *Comment évaluer l’impact des politiques publiques ? Un guide à l’usage des décideurs et praticiens*. [130]

- France Stratégie, R. Desplat and M. Ferracci (n.d.), *Comment évaluer l'impact des politiques publiques ? Un guide à l'usage des décideurs et praticiens*. [134]
- Gaarder, M. and B. Briceño (2010), "Institutionalisation of government evaluation: balancing trade-offs", *Journal of Development Effectiveness*, Vol. 2/3, pp. 289-309, <http://dx.doi.org/10.1080/19439342.2010.505027>. [55]
- Gasper, D. (2018), "Policy Evaluation: From Managerialism and Econocracy to a Governance Perspective", in *International Development Governance*, Routledge, <http://dx.doi.org/10.4324/9781315092577-37>. [28]
- Gasper, D. (2018), "Policy Evaluation: From Managerialism and Econocracy to a Governance Perspective", in *International Development Governance*, Routledge, <http://dx.doi.org/10.4324/9781315092577-37>. [30]
- Gauthier, B. (2015), "Some pointers concerning Evaluation Utilization". [168]
- Goldstein, C. et al. (2018), *Ethical issues in pragmatic randomized controlled trials: A review of the recent literature identifies gaps in ethical argumentation*, BioMed Central Ltd., <http://dx.doi.org/10.1186/s12910-018-0253-x>. [51]
- Government Policy Analysis Unit (2017), *Global Evidence Policy Units: Finland*, https://www.ksi-indonesia.org/file_upload/Evidence-Policy-Unit-in-Finland-the-Government-Po-14Jun2017163532.pdf. [90]
- Greenhalgh, T. et al. (2004), "Diffusion of Innovations in Service Organizations: Systematic Review and Recommendations", *The Milbank Quarterly*, Vol. 82/4, pp. 581-629, <http://dx.doi.org/10.1111/j.0887-378X.2004.00325.x>. [110]
- Haynes, A. et al. (2012), "Identifying Trustworthy Experts: How Do Policymakers Find and Assess Public Health Researchers Worth Consulting or Collaborating With?", *PLoS ONE*, Vol. 7/3, p. e32665, <http://dx.doi.org/10.1371/journal.pone.0032665>. [170]
- Haynes, A. et al. (2018), "What can we learn from interventions that aim to increase policy-makers' capacity to use research? A realist scoping review", *Health Research Policy and Systems*, Vol. 16/1, p. 31, <http://dx.doi.org/10.1186/s12961-018-0277-1>. [173]
- Heider, C. (2017), *Rethinking Evaluation - Efficiency, Efficiency, Efficiency*, <https://ieg.worldbankgroup.org/blog/rethinking-evaluation-efficiency>. [29]
- Hildén, M. (2014), "Evaluation, assessment, and policy innovation: exploring the links in relation to emissions trading", *Environmental Politics*, Vol. 23/5, pp. 839-859, <http://dx.doi.org/10.1080/09644016.2014.924199>. [41]
- HM Treasury (2011), *The Magenta Book: Guidance for evaluation*, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/220542/magenta_book_combined.pdf (accessed on 18 June 2019). [9]
- HM Treasury (2011), *The Magenta Book: Guidance for evaluation*, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/220542/magenta_book_combined.pdf (accessed on 18 June 2019). [191]
- Howlett, M. (2019), "Policy analytical capacity and evidence-based policy-making: Lessons from Canada", https://doi.org/10.1111/j.1754-7121.2009.00070_1.x. [3]

- IGEES (2014), *Irish Government Economic and Evaluation Service*, <https://igees.gov.ie/> [153]
(accessed on 28 January 2019).
- Impact Assessment Office (2018), “The Uncompleted Evaluation of Legislative in Italy: Critical Issues, Prospects and Good Practice”, [94]
<http://www.senato.it/service/PDF/PDFServer/BGT/01082854.pdf> (accessed on 23 September 2019).
- Independent Evaluation Office of UNDP (2019), *UNDP Evaluation Guidelines*. [127]
- Innovate UK (2018), “Evaluation Framework, How we assess our impact on business and the economy”, [82]
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/681741/17.3253_Innovate_UK_Evaluation_Framework_RatherNiceDesign_V2_FINAL_WEB.pdf (accessed on 20 September 2019).
- International Organisation of Supreme Audit Institutions (2019), *ISSAI 100: Fundamental Principles of Public-Sector Auditing*, <http://www.issai.org><http://www.intosai.org> (accessed on 8 January 2020). [20]
- INTOSAI (2016), *Guidelines on the Evaluation of Public Policies*, <http://www.issai.org> [155]
(accessed on 22 August 2019).
- INTOSAI (2010), *Program Evaluation for SAIs, A Primer*, [122]
<https://www.eurosai.org/handle404?exporturi=/export/sites/eurosai/content/documents/materials/Program-Evaluation-for-SAIs.pdf> (accessed on 22 August 2019).
- INTOSAI Working Group on Evaluation of Public Policies and Programs (2019), [188]
Implementation of the INTOSAI GOV 9400 Guidelines: Survey Results.
- INTOSAI and Cour des Comptes (France) (2019), *Implementation of the INTOSAI GOV 9400 Guidelines: Survey results*. [97]
- Jacob, S. and Y. Boisvert (2010), *To Be or Not to Be a Profession: Pros, Cons and Challenges for Evaluation*, <http://dx.doi.org/10.1177/1356389010380001>. [118]
- Jacob, S., S. Speer and J. Furubo (2015), “The institutionalization of evaluation matters: Updating the International Atlas of Evaluation 10 years later”, *Evaluation*, Vol. 21/1, pp. 6-31, <http://dx.doi.org/10.1177/1356389014564248>. [63]
- Johnson, K. et al. (2009), “Research on Evaluation Use A Review of the Empirical Literature From 1986 to 2005”, <http://dx.doi.org/10.1177/1098214009341660>. [167]
- King, J. et al. (2001), *Toward a Taxonomy of Essential Evaluator Competencies*. [149]
- Kothari, A. et al. (2009), “Is research working for you? validating a tool to examine the capacity of health organizations to use research”, *Implementation Science*, Vol. 4/1, p. 46, <http://dx.doi.org/10.1186/1748-5908-4-46>. [181]
- Kroll, A. and D. Moynihan (2018), “The Design and Practice of Integrating Evidence: Connecting Performance Management with Program Evaluation”, *Public Administration Review*, Vol. 78/2, pp. 183-194, <http://dx.doi.org/10.1111/puar.12865>. [17]

- Kusters, C. (2011), *Making evaluations matter: a practical guide for evaluators MSP book View project Strengthening Managing for Impact (SMIP) View project*, <https://www.researchgate.net/publication/254840956>. [103]
- Kusters, C. et al. (2011), "Making evaluations matter: a practical guide for evaluators", Centre for Development Innovation, Wageningen University & Research centre., <https://www.researchgate.net/publication/254840956>. [125]
- Langer, L., J. Tripney and D. Gough (2016), *The science of using science: researching the use of Research evidence in decision-making..* [175]
- Lazaro, B. (2015), *Comparative Study on the Institutionalization of Evaluation in Europe and Latin America*, Eurosocal Programme. [5]
- Lázaro, B. (2015), *Comparative study on the institutionalisation of evaluation in Europe and Latin America*, Eurosocal Programme, Madrid, http://sia.eurosocial-ii.eu/files/docs/1456851768-E_15_ENfin.pdf (accessed on 9 July 2019). [54]
- Ledermann, S. (2012), "Exploring the Necessary Conditions for Evaluation Use in Program Change", *American Journal of Evaluation*, Vol. 33/2, pp. 159-178, <http://dx.doi.org/10.1177/1098214011411573>. [57]
- Ledermann, S. (2012), "Exploring the Necessary Conditions for Evaluation Use in Program Change", <http://dx.doi.org/10.1177/1098214011411573>. [163]
- Leviton, L. and E. Hughes (1981), *Research on the Utilization of Evaluations: A Review and Synthesis view and Synthesis.* [100]
- Little, B. (ed.) (1979), *Speaking Truth to Power: The Art and Craft of Policy Analysis.* [135]
- Liverani, M., B. Hawkins and J. Parkhurst (2013), *Political and institutional influences on the use of evidence in public health policy. A systematic review.*, <http://dx.doi.org/10.1371/journal.pone.0077404>. [68]
- Mackay, K. (2007), *How to Build M&E Systems to Support Better Government*, The World Bank, <http://dx.doi.org/10.1596/978-0-8213-7191-6>. [58]
- Maeda, A., M. Harrit and S. Mabuchi (2012), *Human Development Creating Evidence for Better Health Financing Decisions A Strategic Guide for the Institutionalization of National Health Accounts*, The World Bank, Washington, DC, <http://dx.doi.org/10.1596/978-0-8213-9469-4>. [69]
- Malčík, M. and A. Seberová (2010), "Meta-evaluation and Quality Standard of Final Evaluation Report. The New Educational Review", *The New Educational Review*, Vol. 22, pp. 149-164. [140]
- McDavid, J., I. Huse and L. Hawthorn (2006), *Program Evaluation and Performance Measurement: An Introduction to Practice*, <https://study.sagepub.com/mcdavid3e> (accessed on 28 January 2020). [10]
- Mcguire, M. and R. Zorzi (2005), "EVALUATOR COMPETENCIES AND PERFORMANCE DEVELOPMENT", *The Canadian Journal of Program Evaluation*, Vol. 20/2, pp. 73-99. [146]

- Mergaert, L. and R. Minto (2015), “Ex Ante and Ex Post Evaluations: Two Sides of the Same Coin? The Case of Gender Mainstreaming in EU Research Policy”, *Symposium on Policy Evaluation in the EU*, <http://dx.doi.org/10.1017/S1867299X0000427X>. [37]
- Mideplan (2018), *Guide for Terms of References (Guía de Términos de Referencia)*, <https://documentos.mideplan.go.cr/share/s/DVyxct0OR3a0T6E2QbfQww> (accessed on 21 August 2019). [126]
- Mideplan (2018), *Guide for the use of evaluations: guidelines for its implementation and follow-up on recommendations*, <https://documentos.mideplan.go.cr/share/s/DDVZ114kTjCsTAxihi5Kw> (accessed on 3 September 2019). [180]
- Mideplan (2018), *National Evaluation Policy*, <https://documentos.mideplan.go.cr/share/s/Ymx1WmMJTOWe9YyjeCHKQ> (accessed on 20 September 2019). [83]
- Ministerio de Desarrollo Social y Familia de Chile (2019), *Evaluación Social Ex Ante*, <http://sni.ministeriodesarrollosocial.gob.cl/evaluacion-iniciativas-de-inversion/evaluacion-ex-ante/> (accessed on 5 July 2019). [35]
- Ministerio de Planificación Nacional y Política Económica (2018), *Manual de Evaluación para Intervenciones Públicas*, <https://documentos.mideplan.go.cr/share/s/6eepeLCEsrKkft6Mf5SToA> (accessed on 5 August 2019). [24]
- Ministry of Finance (2006), *The Norwegian Government Agency for Financial Management (DFØ)*, <https://www.regjeringen.no/en/dep/fin/about-the-ministry/etater-og-virksomheter-under-finansdepartementet/subordinateagencies/the-norwegian-government-agency-for-fina/id270409/> (accessed on 12 July 2019). [38]
- Ministry of Finance (Lithuania) (2011), *Recommendations on Implementation of Programs Evaluation Methodology*, https://finmin.lrv.lt/uploads/finmin/documents/files/LT_ver/Veiklos_sritys/Veiklos_efektyvumo_tobulinimas/PVrekomendacijos2011.pdf (accessed on 27 August 2019). [132]
- Ministry of Finance of The Netherlands (2018), *Arrangements for periodic evaluation research*, <https://wetten.overheid.nl/BWBR0040754/2018-03-27> (accessed on 12 July 2019). [6]
- Ministry of Internal Affairs and Communications (2017), “Basic Guidelines for Implementing Policy Evaluation (Revised)”, https://www.soumu.go.jp/main_content/000556221.pdf (accessed on 16 September 2019). [76]
- Ministry of Internal and Communications (2017), *Basic Guidelines for Implementing Policy Evaluation Revised*, http://www.soumu.go.jp/main_content/000556221.pdf. [190]
- Morton, M. (2009), *Applicability of Impact Evaluation to Cohesion Policy 1 Report Working Paper of*, https://ec.europa.eu/regional_policy/archive/policy/future/pdf/4_morton_final-formatted.pdf (accessed on 8 August 2019). [49]
- National Audit Office (GBR) (2013), *Evaluation in Government*, <https://www.nao.org.uk/report/evaluation-government/> (accessed on 22 August 2019). [154]

- Neuhoff, A. et al. (2015), *The What Works Marketplace Helping Leaders Use Evidence to Make Smarter Choices 2 Invest in What Works Policy Series*, The Bridgespan group, <http://www.results4america.org>. [177]
- Newman, J., A. Cherney and B. Head (2017), "Policy capacity and evidence-based policy in the public service", *Public Management Review*, Vol. 19/2, pp. 157-174, <http://dx.doi.org/10.1080/14719037.2016.1148191>. [164]
- Newman, K., C. Fisher and L. Shaxson (2012), "Stimulating Demand for Research Evidence: What Role for Capacity-building?", *IDS Bulletin*, Vol. 43/5, pp. 17-24, <http://dx.doi.org/10.1111/j.1759-5436.2012.00358.x>. [111]
- Newman, K., C. Fisher and L. Shaxson (2012), "Stimulating Demand for Research Evidence: What Role for Capacity-building?", *IDS Bulletin*, Vol. 43/5, pp. 17-24, <http://dx.doi.org/10.1111/j.1759-5436.2012.00358.x>. [183]
- OECD (2020), *Building capacity for evidence informed policy making*, OECD, Paris. [99]
- OECD (2020), *The Irish Government Economic and Evaluation Service: Using Evidence-Informed Policy Making to Improve Performance*, OECD Public Governance Reviews, OECD Publishing, Paris, <https://dx.doi.org/10.1787/cdda3cb0-en>. [98]
- OECD (2019), *Budgeting and Public Expenditures in OECD Countries 2019*, OECD Publishing, Paris, <https://dx.doi.org/10.1787/9789264307957-en>. [12]
- OECD (2019), *Evaluating Public Sector Innovation Support or hindrance to innovation?*, Observatory of Public Sector Innovation-OPSI, Paris, <https://oecd-opsi.org/wp-content/uploads/2019/05/Evaluating-Public-Sector-Innovation-Part-5a-of-Lifecycle-Report.pdf> (accessed on 11 September 2019). [59]
- OECD (2019), "Open Government Data Report". [107]
- OECD (2019), *Open Government in Biscay*, OECD Public Governance Reviews, OECD Publishing, Paris, <https://dx.doi.org/10.1787/e4e1a40c-en>. [11]
- OECD (2018), *Building Capacity for Evidence Informed Policy Making: Towards a Baseline Skill Set*, <http://www.oecd.org/gov/building-capacity-for-evidence-informed-policy-making.pdf> (accessed on 3 September 2019). [1]
- OECD (2018), *Centre Stage 2- The organisation and functions of the centre of government in OECD countries*, <https://www.oecd.org/gov/centre-stage-2.pdf>. [93]
- OECD (2018), *Cost-Benefit Analysis and the Environment: Further Developments and Policy Use*, OECD Publishing, Paris, <https://dx.doi.org/10.1787/9789264085169-en>. [47]
- OECD (2018), *Draft Policy Framework on Sound Public Governance*, <http://www.oecd.org/gov/draft-policy-framework-on-sound-public-governance.pdf> (accessed on 8 July 2019). [2]
- OECD (2018), "OECD Best Practice Principles for Regulatory Policy: Reviewing the Stock of Regulation". [123]
- OECD (2018), *OECD Performance Budgeting Survey*. [192]

- OECD (2018), *OECD Regulatory Policy Outlook 2018*, OECD Publishing, Paris, [172]
<https://dx.doi.org/10.1787/9789264303072-en>.
- OECD (2017), *Governing Better Through Evidence-Informed Policy Making Options for an OECD Work Agenda*. [112]
- OECD (2017), *Government at a Glance - OECD*, OECD, Paris, [4]
<http://www.oecd.org/gov/govataglance.htm> (accessed on 9 July 2019).
- OECD (2017), *Making policy evaluation work*. [105]
- OECD (2017), *Policy Advisory Systems: Supporting Good Governance and Sound Public Decision Making*, OECD Public Governance Reviews, OECD Publishing, Paris, [186]
<https://dx.doi.org/10.1787/9789264283664-en>.
- OECD (2017), *Systems Approaches to Public Sector Challenges: Working with Change*, OECD Publishing, Paris, [53]
<https://dx.doi.org/10.1787/9789264279865-en>.
- OECD (2016), *Engaging Public Employees for a High-Performing Civil Service*, OECD Public Governance Reviews, OECD Publishing, Paris, [184]
<https://dx.doi.org/10.1787/9789264267190-en>.
- OECD (2016), *Evaluation Systems in Development Co-operation: 2016 Review*, OECD Publishing, Paris, [144]
<https://dx.doi.org/10.1787/9789264262065-en>.
- OECD (2016), *Open Government: The Global Context and the Way Forward*, OECD Publishing, Paris, [43]
<https://dx.doi.org/10.1787/9789264268104-en>.
- OECD (2016), *Supreme Audit Institutions and Good Governance: Oversight, Insight and Foresight*, OECD Public Governance Reviews, OECD Publishing, Paris, [21]
<https://doi.org/10.1787/9789264263871-en>.
- OECD (2014), "Budget Review: Germany", *OECD Journal on Budgeting*, Vol. 2. [81]
- OECD (2014), "Centre Stage Driving Better Policies from the Centre of Government", [92]
[http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=gov/pgc/mpm\(2014\)3&doclanguage=en](http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=gov/pgc/mpm(2014)3&doclanguage=en) (accessed on 23 September 2019).
- OECD (2014), *Chile's Supreme Audit Institution: Enhancing Strategic Agility and Public Trust*, OECD Public Governance Reviews, OECD Publishing, Paris, [96]
<https://dx.doi.org/10.1787/9789264207561-en>.
- OECD (2011), "Government at a Glance". [174]
- OECD (2011), "Typology and implementation of spending reviews", *OECD SBO Meeting on Performance and Results*, [13]
[http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=GOV/PGC/SBO\(2011\)9&doclanguage=en](http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=GOV/PGC/SBO(2011)9&doclanguage=en) (accessed on 1 August 2019).
- OECD (2010), "DAC Guidelines and Reference Series: Quality Standards for Development Evaluation", <https://www.oecd.org/development/evaluation/qualitystandards.pdf> (accessed on 9 July 2019). [48]
- OECD (2010), *Good Practices in Supporting Supreme Audit Institutions*, [95]
<http://www.oecd.org/dac/effectiveness/Final%20SAI%20Good%20Practice%20Note.pdf>.

- OECD (2008), *Performance Budgeting: A Users' Guide*, [19]
<https://www.oecd.org/gov/budgeting/Performance-Budgeting-Guide.pdf> (accessed on 2 August 2019).
- OECD (forthcoming), *Ensuring the Good Governance of Evidence, taking stock of standards for policy design, implementation and evaluation*, OECD, Paris. [106]
- OECD-DAC (2009), "Guidelines for Project and Programme Evaluations", [84]
https://www.entwicklung.at/fileadmin/user_upload/Dokumente/Projektentwicklung/Englisch/Guidelines_for_Project_and_Programme_Evaluations.PDF (accessed on 20 September 2019).
- OECD-DAC (2002), *Glossary of Key Terms in Evaluation and Results Based Management*, [26]
<https://www.oecd.org/dac/evaluation/18074294.pdf> (accessed on 18 June 2019).
- Office fédéral de la justice (2005), *Guide for the evaluation of the efficacy of the Confederation (Guide de l'évaluation de l'efficacité à la Confédération)*, [137]
http://www.ofj.admin.ch/ejpd/fr/home/themen/staat_und_buerger/ref_evaluation/ref_umsetzung_art.html (accessed on 27 August 2019).
- Office of Management and Budget (2020), *MB-20-12 Program Evaluation Standards and Practices*, [86]
<https://www.whitehouse.gov/wp-content/uploads/2020/03/M-20-12.pdf>.
- Office of Management and Budget (2018), *Monitoring and Evaluation Guidelines for Federal Departments and Agencies that Administer United States Foreign Assistance*, [136]
<https://www.whitehouse.gov/wp-content/uploads/2017/11/M-18-04-Final.pdf> (accessed on 27 August 2019).
- Olejniczak, K. and E. Raimondo (2016), "Evaluation units as knowledge brokers: Testing and calibrating an innovative framework", *Evaluation*, Vol. 22/2, pp. 168-189, [182]
<http://dx.doi.org/10.1177/1356389016638752>.
- Olejniczak, K., E. Raimondo and T. Kupiec (2016), "Evaluation units as knowledge brokers: Testing and calibrating an innovative framework", *Evaluation*, Vol. 22/2, pp. 168-189, [64]
<http://dx.doi.org/10.1177/1356389016638752>.
- Oliver, K. et al. (2015), "Identifying public health policymakers' sources of information: comparing survey and network analyses", *The European Journal of Public Health*, Vol. 27/suppl_2, p. ckv083, [169]
<http://dx.doi.org/10.1093/eurpub/ckv083>.
- OMB (2010), *Section 200 - Overview of the Federal Performance Framework*. [178]
- Operational and Evaluation Audit Division of Costa Rica (2014), *INFORME DE AUDITORÍA DE CARÁCTER ESPECIAL SOBRE LOS PROCESOS DE SEGUIMIENTO, EVALUACIÓN Y RENDICIÓN DE CUENTAS PÚBLICA EN*, [156]
https://cgrfiles.cgr.go.cr/publico/jaguar/sad_docs/2015/DFOE-SAF-IF-09-2014-Recurrido.pdf (accessed on 22 August 2019).
- Parkhurst, J. (2017), *The politics of evidence : from evidence-based policy to the good governance of evidence*, Routledge, London, <http://researchonline.lshtm.ac.uk/3298900/> (accessed on 23 November 2018). [32]
- Patton, M. (1978), "Utilization-focused evaluation". [102]

- Picciotto, R. (2013), "Evaluation Independence in Organizations", *Journal of MultiDisciplinary Evaluation*, Vol. 9/20, p. 15. [56]
- Picciotto, R. (n.d.), *The Value of Evaluation Standards: A Comparative Assessment*, [120]
<http://evaluation.wmich.edu/jmde/Articles>.
- Pleger, L. and S. Hadorn (2018), "The big bad wolf's view: The evaluation clients' perspectives on independence of evaluations", *Evaluation*, Vol. 24/4, pp. 456-474, [115]
<http://dx.doi.org/10.1177/1356389018796004>.
- Pleger, L. and S. Hadorn (2018), "The big bad wolf's view: The evaluation clients' perspectives on independence of evaluations", *Evaluation*, Vol. 24/4, pp. 456-474, [138]
<http://dx.doi.org/10.1177/1356389018796004>.
- Podems, D. (2013), "Evaluator competencies and professionalizing the field: Where are we now?", *Canadian Journal of Program Evaluation*, Vol. 28/3, pp. 127-136. [150]
- Poder Ejecutivo Nacional (2018), *Decreto 292/2018: Evaluación de Políticas y Programas Sociales*, 11-04-2018, [33]
<http://servicios.infoleg.gob.ar/infolegInternet/verNorma.do?id=308653> (accessed on 8 July 2019).
- Poder Ejecutivo Nacional de Argentina (2018), *Decreto 292/2018: Evaluación de Políticas y Programas Sociales*. [22]
- Polish Ministry of Infrastructure and Development (2015), *Guidelines of cohesion policy evaluation for period 2014-2020*, [145]
http://www.ewaluacja.gov.pl/media/13209/wytyczne_090915_final.pdf (accessed on 22 August 2019).
- Punton, M. et al. (2016), *HOW CAN CAPACITY DEVELOPMENT PROMOTE EVIDENCE-INFORMED POLICY MAKING? Literature Review for the Building Capacity to Use Research Evidence (BCURE) Programme*, <http://www.itad.com/knowledge-and-resources/bcure> (accessed on 6 September 2019). [185]
- Results for America (2017), *Government Mechanisms to Advance the Use of Data and Evidence in Policymaking: A Landscape Review*. [166]
- Results for America (2017), *Government Mechanisms to Advance the Use of Data and Evidence in Policymaking: A Landscape Review*. [193]
- Robert Picciotto (2013), "Evaluation Independence in Organizations", *Journal of MultiDisciplinary Evaluation*, Vol. 9/20, p. 15. [114]
- Robinson, M. (2014), *Spending reviews*, <http://www.pfmresults.com>. (accessed on 25 June 2019). [187]
- Roh, J. (2018), "Improving the government performance management system in South Korea", *Asian Education and Development Studies*, Vol. 7/3, pp. 266-278, [79]
<http://dx.doi.org/10.1108/AEDS-11-2017-0112>.
- Rutter, J. (2012), *Evidence and Evaluation in Policy making*, Institute for Government, [52]
https://www.instituteforgovernment.org.uk/sites/default/files/publications/evidence%20and%20evaluation%20in%20template_final_0.pdf.

- Schillemans, T. and M. Bovens (2011), *The Challenge of Multiple Accountability: Does Redundancy lead to Overload?*, [62]
- Schoenefeld, J. and A. Jordan (2017), "Governing policy evaluation? Towards a new typology", *Evaluation*, Vol. 23/3, pp. 274-293, [39]
<http://dx.doi.org/10.1177/1356389017715366>.
- Scriven, M. (1969), "An introduction to meta-evaluation", *Educational Product Report*, Vol. 2. [139]
- Secretaria de Desarrollo Social (2015), *Decree for which the Council of social Development Policy Evalaution is regulated.*, [91]
- Shaxson, L. (2019), "Uncovering the practices of evidence-informed policy-making", *Public Money & Management*, Vol. 39/1, pp. 46-55, [108]
<http://dx.doi.org/10.1080/09540962.2019.1537705>.
- Sinatra, G., D. Kienhues and B. Hofer (2014), "Educational Psychologist Addressing Challenges to Public Understanding of Science: Epistemic Cognition, Motivated Reasoning, and Conceptual Change", <http://dx.doi.org/10.1080/00461520.2014.916216>. [113]
- Smismans, S. (2015), "Policy Evaluation in the EU: The Challenges of Linking Ex Ante and Ex Post Appraisal", *Symposium on Policy Evaluation in the EU*, [14]
<http://dx.doi.org/10.1017/S1867299X00004244>.
- Smismans, S. (2015), "Policy Evaluation in the EU: The Challenges of Linking Ex Ante and Ex Post Appraisal", *Symposium on Policy Evaluation in the EU*, [34]
<http://dx.doi.org/10.1017/S1867299X00004244>.
- Stern, E., M. Saunders and N. Stame (2015), "Standing back and looking forward: Editors' reflections on the 20th Anniversary of Evaluation", *Evaluation*, Vol. 21/4, pp. 380-390, [101]
<http://dx.doi.org/10.1177/1356389015608757>.
- Steuerle, E. and L. Jackson (2016), *Advancing the power of economic evidence to inform investments in children, youth, and families*, National Academies Press, [46]
<http://dx.doi.org/10.17226/23481>.
- Stevahn, L. et al. (2005), "Establishing Essential Competencies for Program Evaluators", *ARTICLE American Journal of Evaluation*, <http://dx.doi.org/10.1177/1098214004273180>. [147]
- Stufflebeam, D. (2001), *Method Notes Evaluation Checklists: Practical Tools for Guiding and Judging Evaluations*, <http://www.wmich.edu/evalctr/checklists/>. [143]
- Stufflebeam, D. (1978), "Meta evaluation: an overview", *Evaluation and The Health Professions*, Vol. 1/1, <https://journals.sagepub.com/doi/pdf/10.1177/016327877800100102> [141]
 (accessed on 19 August 2019).
- Superu (2018), *Making sense of evidence: A guide to using evidence in policy*, [179]
<https://thehub.sia.govt.nz/assets/Uploads/Making-Sense-of-Evidence-handbook-FINAL.pdf>
 (accessed on 3 September 2019).
- Swiss Federal Audit Office (2019), *Involving Stakeholders in Evaluation at the Swiss Federal Audit Office*, http://www.program-evaluation.ccomptes.fr/images/stories/evenements/Vilnius_2019/Presentation_Switzerland_Stakeholder_Involvement_WGEPPP_2019.pdf [189]
 (accessed on 6 September 2019).

- The Cabinet Secretariat (2019), *The status-quo about the promotion of statistics reform*, [74]
<http://www.kantei.go.jp/jp/singi/toukeikaikaku/dai5/siryou1.pdf> (accessed on 2 September 2019).
- The Committee on Promoting EBPM (2017), *Guidelines on securing and developing human resources for the promotion of EBPM*, [73]
<https://www.gyokaku.go.jp/ebpm/img/guideline1.pdf> (accessed on 2 September 2019).
- The European Network for Rural Development (2014), *The Ex Ante Evaluation 2014-2020 RDPs*, [36]
https://enrd.ec.europa.eu/evaluation/publications/guidelines-ex-ante-evaluation-2014-2020-rdps_en (accessed on 29 January 2020).
- The Ministry of Internal Affairs and Communication (2005), *Policy Evaluation Implementation Guidelines*, [78]
https://www.soumu.go.jp/main_content/000556222.pdf.
- The Ministry of Internal Affairs and Communications (2010), "Guidelines for Publication of Information on Policy Evaluation", [77]
http://www.soumu.go.jp/main_content/000556224.pdf (accessed on 16 September 2019).
- The Statistical Reform Promotion Council (2017), *The final report of the Statistical Reform Promotion Council. (In Japanese)*, [72]
http://www.kantei.go.jp/jp/singi/toukeikaikaku/pdf/saishu_honbun.pdf (accessed on 2 September 2019).
- The World Bank (2018), *Spending Review Manual: Bulgaria*. [15]
- Treasury Board Secretariat (2019), *Evaluation in the Government of Canada - Canada.ca*, [87]
<https://www.canada.ca/en/treasury-board-secretariat/services/audit-evaluation/evaluation-government-canada.html> (accessed on 20 September 2019).
- Treasury Board Secretariat (2013), *Assessing program resource utilization when evaluating federal programs*, [88]
<https://www.canada.ca/en/treasury-board-secretariat/services/audit-evaluation/centre-excellence-evaluation/assessing-program-resource-utilization-evaluating-federal-programs.html> (accessed on 2 August 2019).
- Treasury Board Secretariat (2010), *Supporting Effective Evaluations: A Guide to Developing Performance Measurement Strategies*, [89]
<https://www.canada.ca/en/treasury-board-secretariat/services/audit-evaluation/centre-excellence-evaluation/guide-developing-performance-measurement-strategies.html> (accessed on 2 August 2019).
- United Nations Evaluation Group (2016), *Norms and Standards for Evaluation*. [121]
- United States Office of Management and Budget (2019), *Phase 1 Implementation of the Foundations for Evidence-Based Policymaking Act of 2018: Learning Agendas, Personnel, and Planning Guidance*, [75]
<https://www.whitehouse.gov/wp-content/uploads/2019/07/M-19-23.pdf>.
- Vaessen, J. (2018), *New blogpost - Five ways to think about quality in evaluation*, [104]
<https://www.linkedin.com/pulse/new-blogpost-five-ways-think-quality-evaluation-jos-vaessen> (accessed on 21 June 2019).
- Vammalle, C. and A. Ruiz Rivadeneira (2017), "Budgeting in Chile", *OECD Journal on Budgeting*, Vol. 16/3, [44]
<https://dx.doi.org/10.1787/budget-16-5jfw22b3c0r3>.

- van Ooijen, C., B. Ubaldi and B. Welby (2019), "A data-driven public sector: Enabling the strategic use of data for productive, inclusive and trustworthy governance", *OECD Working Papers on Public Governance*, No. 33, OECD Publishing, Paris, <https://dx.doi.org/10.1787/09ab162c-en>. [71]
- Viñuela, L., D. Ortega and F. Gomes (2015), *Technical Note - Mechanisms and incentives for the adoption of evaluation of Policies and Programs to improve the Efficiency of Public Expenditure*. [66]
- Walker, K. and K. Moore (2011), *Performance Management and Evaluation: What's the difference?*, <https://www.childtrends.org/wp-content/uploads/2013/06/2011-02PerformMgmt.pdf> (accessed on 18 July 2019). [16]
- Weiss, C. (1993), "Where politics and evaluation research meet", *Evaluation Practice*, Vol. 14/1, pp. 93-106, [http://dx.doi.org/10.1016/0886-1633\(93\)90046-R](http://dx.doi.org/10.1016/0886-1633(93)90046-R). [40]
- Weiss, C. and C. Weiaa Harvard (1998), "Have We Learned Anything New About the Use of Evaluation?", *American Journal of Evaluation*, Vol. 19/1, pp. 21-33. [160]
- World Bank et al. (2019), *World Bank Group Evaluation Principles*, <http://www.worldbank.org>. [124]
- Yang, S. and A. Torneo (2016), "Government Performance Management and Evaluation in South Korea: History and Current Practices", *Public Performance & Management Review*, Vol. 39/2, pp. 279-296, <http://dx.doi.org/10.1080/15309576.2015.1108767>. [80]
- Zida, A. et al. (2017), "Evaluating the Process and Extent of Institutionalization: A Case Study of a Rapid Response Unit for Health Policy in Burkina Faso.", *International journal of health policy and management*, Vol. 7/1, pp. 15-26, <http://dx.doi.org/10.15171/ijhpm.2017.39>. [70]

Annex 2.A. Annex guidelines and methods

Table 2.4. Guidelines and methods for policy evaluation

Country	Year	Author	Title
Australia	2014	Department of Finance	Resource Management Guidance for the Public Governance, Performance and Accountability Act 2013 RMG 131 Developing Good Performance Information
	2013	Department of the Prime Minister and Cabinet	Monitoring, review and evaluation (Cabinet Implementation Unit Toolkit)
Austria	2013	Federal Chancellery	Handbook for Performance Management
	2013	Federal Chancellery	Handbook for Performance Management
Canada	2017	Treasury Board Secretariat	Guide to Rapid Impact Evaluation
	2013	Treasury Board Secretariat	Assessing Program Resource Utilization When Evaluating Federal Programs
	2012	Treasury Board Secretariat	Theory-Based Approaches to Evaluation: Concepts and Practices
	2010	Treasury Board Secretariat	Supporting Effective Evaluations: A Guide to Developing Performance Measurement Strategies
	2019	Treasury Board Secretariat	Integrating Gender-Based Analysis Plus into Evaluation: A Primer
Czech Republic	2016	Min.Regional Development	Methodological guidance for evaluations in the 2014-2020 programming period
Estonia	2012	Government of Estonia	Methodology of Impact Assessment
	2012	Estonian Evaluation Association	Good Public Evaluation Code of Practice
	2011	Government of Estonia	Good Public Engagement Code of Practice
Finland	Annually	Council of State/PMO	Government's Annual Plan for research, foresight and evaluation
France	2017	France Stratégie	Guide de l'évaluation socio-économique des investissements publics
	2016	France Stratégie	Comment évaluer l'impact des politiques publiques : un guide l'usage des décideurs et des praticiens
	2010	INSEE	Méthodes économétriques pour l'évaluation des politiques publiques
Germany		Youth Experimentation Fund	Guide méthodologique relative aux évaluations du FEJ
	2016	Federal Ministry of the Interior	Handbuch für Organisationsuntersuchungen und Personalbedarfsermittlung
	2011	Federal Ministry of Finance	Arbeitsanleitung Einführung in Wirtschaftlichkeitsuntersuchungen
	2007	Federal Ministry of the Interior	Empfehlungen für interne Revisionen in der Bundesverwaltung
	2000	Federal Ministry for Family Affairs	Zielgeführte Evaluation von Programmen
Great Britain	2018	HM Treasury	Guide to developing the project business case
	2018	HM Treasury	Guide to developing the programme business case
	2018	Better Regulation Executive	Better regulation framework
	2015	Government Social Research Service (HM Treasury)	Government Social Research Publication Protocol
Greece	2018	Secretariat General of the Government	Manual of Inter-Ministerial Coordination
	2015	European Commission	Better Regulation Toolbox
	2015	European Commission	Better Regulation Guidelines

Country	Year	Author	Title
Ireland	2018	Department of Public Expenditure & Reform	Public Spending Code
Italy	2018	PCM	Guidelines RIA
	2017	NUVAP	Guidelines for ex post and ongoing evaluations: requesting and using evaluations
	2017	NUVAP	GL ex post & ongoing evaluations
	2017	Decree President Council of Ministers	Guidelines for ex-ante and ex-post impact analysis of regulatory acts
	2015	NUVAP	GL Evaluation Plans
	2015	NUVAP	Evaluation Plans 2014-2020: general orientation and a short guide on available guidance
Japan	2017	Ministry of Internal Affairs and Communications	Basic Guidelines for Implementing Policy Evaluation (Revised)
	2013	Ministry of Internal Affairs and Communications	Target Management-based Policy Evaluation Implementation Guidelines
	2010	Ministry of Internal Affairs and Communications	Guidelines for Publication of Information on Policy Evaluation
	2010	Ministry of Internal Affairs and Communications	Implementation Guidelines for Policy Evaluation Pertaining to Special Taxation Measures
	2007	Ministry of Internal Affairs and Communications	Implementation Guidelines for Policy Evaluation of Regulations
	2005	Ministry of Internal Affairs and Communications	Policy Evaluation Implementation Guidelines
Korea	2017	Office for Government Policy Coordination	Government Performance Evaluation Manual
Lithuania	2011	Ministry of Finance	Recommendations on Implementation of Programs Evaluation Methodology
	2010	Ministry of Finance	Evaluation of EU structural assistance: Methodological guidance
Latvia	2018	Ministry of Finance	Instruction on Analysis of the Execution of State Budget
	2016	Cross-sectoral Coordination Centre	Manual on Policy Making
Mexico	2007	Ministry of Finance, Ministry of Public Administration, National Council for the Evaluation of Social Development Policy	General Guidelines for the Evaluation of Federal Programs
Norway	2018	DFO	Strategic and systematic use of evaluation in management/governance
	2009	Ministry of Justice	Evaluation of laws
	2007	DFO	Evaluation of central governmental grants
	2005	Ministry of Finance	Guidelines to carry out evaluations
New Zealand	2018	Superu	Making sense of evidence: A guide to using evidence in policy
	2015	Superu	Evaluation Standards for People Commissioning, Using, Participating in, or Conducting Evaluations
Poland	2018	Ministry of Economic Development	Guidelines for the evaluation of cohesion policy" (updated)
	2015	Ministry of Economic Development	Guidelines for the evaluation of cohesion policy
Portugal	2018	Juris App	Manual
Spain	2015	AEVAL	Practical guide for the design and implementation of public policy evaluations
	2007	AECID (Spanish Agency for International Cooperation for Development)	Spanish Cooperation Evaluation Management Manual
Slovakia	2016	Ministry of Finance	Value for money
	Forthcoming	Ministry of Economy	RIA 2020
Switzerland	2015	Federal office of justice	Planifier une évaluation, en assurer le suivi et en valoriser les résultats

Country	Year	Author	Title
United States	2013	State secretariat for economic affairs	Analyse d'impact de la réglementation - Manuel
	2012	Federal office of justice	Recommandations de l'Office fédéral de la justice pour la formulation des clauses d'évaluation
	2005	Federal office of justice	Guide de l'évaluation de l'efficacité la Confédération
	2018	OMB	A-11 section 200: an overview of the Federal Performance Framework
	2018	OMB	M-18-04: Monitoring and Evaluation guidelines for agencies that administer foreign assistance
	2019	OMB	M-19-23: guidelines for the implementation of the Foundations for Evidence-Based Policymaking Act of 2018
Argentina	2020	OMB	M-20-12 Phase 4 Implementation of the Foundations for Evidence-Based Policymaking Act of 2018: Program Evaluation Standards and Practices
	2018	National Council for the Coordination of Social Policies	Resolución No.310 Lineamientos de MyE
Brazil	2018	Cabinet Office	Resolución No.212/18 Plan Anual de MyE
	2018	Civil House, Ministry of Finance, Ministry of Planning, Ministry of Transparency and Comptroller General	Public Policies Evaluation: practical guide for ex ante analysis
	2017	Ministry of Social Development	"How to promote impact evaluation in social programs"
	2015	Ministry of Transparency and Comptroller General	Methodology Manual for Evaluating Government Programs Execution
Colombia	2014	Federal Court of Auditors	Referential for Governance Evaluation in Public Policies
	2018	Departamento Administrativo de Planeación Nacional	Guide for the evaluation of public policies
Costa Rica	2018	Ministry of National Planning and Economic Policy	Guide for the use of evaluations: guidelines for its implementation and follow-up on recommendations.
	2017	Ministry of National Planning and Economic Policy	Manual of evaluation for public interventions
	2017	Ministry of National Planning and Economic Policy	Guide on the approach of gender equality and human rights in evaluation: guidelines for its incorporation into the evaluation process.
	2017	Ministry of National Planning and Economic Policy	Guide of evaluability: methodological guidelines for the evaluability of public interventions.
Kazakhstan	2017	Ministry of National Economy	State planning system

Source: OECD Survey on Policy Evaluation (2018).

Chapter 3.

How do countries address the challenges of promoting quality and use of evaluations?

Quality and use of evaluations are essential to ensure relevance and impact on policy-making. They are key to promote learning, accountability and effective contribution of evaluation to decision-making tools such as regulation and budgeting. However, achieving both quality and use is widely recognised as the most important challenge faced by policy-makers and practitioners in this area. This is due to a mix of skills and institutional gaps, heterogeneous oversight of evaluation processes, and insufficient mechanisms for quality control and capacity for uptake of evidence. This chapter discusses the external and internal factors that affect the quality and use of policy evaluations, as well as their interlinkages. It examines the various mechanisms put forth by governments in order to promote the good quality and use of policy evaluations, and highlights relevant country practices in this regard.

Key findings

Quality is key to ensure the robustness of policy evaluations, and can be achieved through *quality control* and *quality assurance* processes. These have been put in place by countries through different means:

- Standards for quality play an important role in quality assurance, but are less likely to be embedded in normative instruments such as legal and policy frameworks.
- Quality control mechanisms are much less common across the sample of countries that responded to the OECD Survey, both within and outside of the executive, and may constitute an area of development in order to ensure that evaluation reports and evaluative evidence meet a high quality standard.

While quality is very important and can facilitate use of evaluation, it is not enough to guarantee such use, which remains an important challenge faced by many countries. However, the use of evaluation is crucial to ensure impact and to promote evidence-informed policy-making and learning.

Organisations and institutional mechanisms within the executive play an important role in creating a market place for the use of evaluations. Yet the heterogeneity of country approaches suggests that there is no one size fits all approach and the set up depends on the local political and cultural context.

Countries have recognised the importance of competences for promoting the quality and use of evaluation. Most mechanisms for the development of skills and competences are aimed at evaluators, managers, or senior civil servants, and aim to ensure high *quality evaluations*. Further increasing the competences of policy and decision-makers and increasing capacity for the *use of evaluation*, on the other hand, may increase demand for evaluative evidence.

The role of institutions outside of the executive remains limited, both in the promotion of quality and use of evaluation, aside from their involvement in the budgetary cycle. Parliament can play a role in some countries. Supreme Audit Institutions play an important role in the supply of unbiased evaluations overall in a significant number of countries.

Introduction

Quality and use of evaluations are essential to ensure impact on policy-making, and thus in ensuring that evaluations actually serve as tools for learning, accountability and better decision-making. However, achieving both quality and use is widely recognised as some of the most important challenge faced by policy-makers and practitioners in this area. This is due to a mix of skills gaps, heterogeneous oversight of evaluation processes, and insufficient mechanisms for quality control and capacity for uptake of evidence.

This chapter discusses the quality and use of policy evaluation, as well as their interlinkages. It discusses both the external and internal factors that affect the quality and use of policy evaluations. Finally, the chapter examines the various mechanisms put forth by governments in order to promote the good quality and utilisation of policy evaluations, as well as highlight some interesting country practices in the area.

Quality and use are essential

Quality matters

Not all evaluations are created equal, some deserve to be given more weight in decision-making. In fact, high quality evaluations generate robust and credible results that can be used with confidence. As a result, good quality evaluations enable policies to be improved and are thus a key part of the policy cycle. In particular, quality impact evaluations provide evidence on the outcome of policies, as well as on whether these changes can be attributed to the intervention in question. In this sense, they facilitate *learning* in decision-making and policy design, by providing reliable information on why and how a policy was successful or not, and the underlying causal mechanisms leading to success or failure.

Quality evaluations also have the potential to increase policy *accountability* as they can provide trustworthy evidence on how resources were spent, what benefits were achieved and what the returns were. Good quality evaluations give citizens and stakeholders access to information on whether the efforts carried out by the government, including allocation of financial resources, are producing the expected results (OECD, 2018^[2]). As such, good quality evaluations are fundamental to democratic accountability (HM Treasury, 2011^[9]).

Conversely, poor quality evaluations carry the risk of providing unfit evidence, or evidence that is subject to bias and undue influence. Poor quality evidence also implies that a policy that is ineffective, or even harmful, might either be implemented or continue to be. Finally, opportunities to use public funds more effectively may be missed.

Use is also important

Effective use of evaluations is key to embed them in policy making processes and to generate incentives for the dissemination of evaluation practices. It is a critical source of feedback for generating new policies and developing rationale for government interventions. If evaluations are not used, gaps will remain between what is known to be effective as suggested by evidence and policy, and decision-making in practice. Simply put, evaluations that are not used represent missed opportunities for learning and accountability.

Connections between evidence and policy-making remain elusive (OECD, 2020^[99]): the use of policy evaluation continues to be one of the most important challenges. This is compounded by the fact that the underuse of evaluations may jeopardize the legitimacy of the evaluative exercise in the first place. When decision-makers ignore the results of evaluations, the claim for further analysis is undermined (Leviton and Hughes, 1981^[100]). Unused evaluations may also contribute to an impression of excess supply, whereby quality evidence gets lost in the shuffle.

Underuse also represents a waste of public resources: policy evaluations, whether conducted internally or contracted-out to external stakeholders, require significant public human and financial resources (Stern, Saunders and Stame, 2015^[101]), which will be lost if they lead to no outcomes.

Quality and use are closely interrelated

Quality and use of evaluations are intrinsically linked, thereby increasing their significance for policy-makers. Some academic authors consider use to be a key component of an evaluation's quality (Patton, 1978^[102]) (Kusters, 2011^[103]) (Vaessen, 2018^[104]). From this perspective, the extent to which an evaluation meets the needs of different groups of users dictates its quality. For instance, stakeholder involvement and iterative learning are seen as the foundation for using evaluation, and by implication, quality. Conversely, evaluations that adhere to the quality standard of appropriateness – that is, evaluations that address multiple political considerations, are useful to achieve policy goals and consider the local context – are by very definition more useful to intended users.

In addition, quality should also be conducive to greater potential for use. In fact, insofar as good quality evaluations benefit from greater credibility, both because they are technically rigorous and well governed, they are likely to be given more weight in decision-making. Similarly, unused data are likely to suffer because they are not subject to critical questioning. However, in practice, it is important to recognise that quality may be associated with greater complexity of the results, due to methodological requirements and limits with the use of quantitative methods, which may make the results difficult to read and interpret for a lay audience.

Exogenous factors affecting quality and use of evaluations

Quality and use can be influenced both by policy and internal factors, amenable to policy intervention, as well as by a range of exogenous factors determined by cultural, historical and environmental circumstances.

The extent to which policies can be evaluated

For an evaluation to be of high quality and to be useful for policy-making, the policy or programme should be easily evaluable in the first place, meaning that it should be possible to evaluate it in a credible and reliable manner (OECD, 2010^[48]). Two main factors may affect the degree to which a policy can easily be evaluated:

- the nature and design of the policy or programme itself
- the quality and availability of non-survey specific data.

The nature and design of the policy

Clearly laying out the objectives of a policy and the levers to attain it will facilitate the evaluation (OECD, 2017^[105]). This also implies the original intentions of the programme developers be explicit and open to critical thinking (OECD, forthcoming^[106]). One way to facilitate clear policy objectives is to develop a theory of change and logic model, which can be done either at the stage of policy design, or when developing an evaluation. A theory of change can be defined as a set of interrelated assumptions explaining how and why an intervention is likely to produce outcomes in the target population (OECD, forthcoming^[106]). Developing a theory of change can lead to better policy planning and evaluation because the policy or programme activities are linked to a detailed and plausible understanding of how change actually happens. A logic model sets out the conceptual connections between concepts in the theory of change to show what intervention, at what intensity, delivered to whom and at what intervals would likely produce specified short term, intermediate and long term outcomes (OECD, forthcoming^[106]).

Box 3.1. Benefits of developing an intervention theory of change and logic model for policy or programme development

The evaluability of the programme —for both implementation and outcomes— is facilitated, by signposting appropriate metrics.

The original intentions of the programme developers are clearly set out, and are explicit and open to critique.

The underlying logic of the assumptions made in the theory, for example, that undertaking a certain activity will lead to a particular outcome, can be scrutinised.

The realism of the assumptions made by the programme developers can be checked against wider evidence of ‘what works’ to assess the likelihood of the programme being successful.

Commissioners can check whether the programme meets their needs; and providers and practitioners delivering the programme can check their own assumptions and the alignment of their expectations against the original intentions of the programme developers.

The key parameters or boundaries (e.g., who is the programme for, and under what specific circumstances) can be set out, reducing the likelihood that the programme is used inappropriately or ineffectively.

Core components (of content, or of implementation, or both) that are believed to be essential to the programme’s effectiveness can be identified.

Activity traps can be identified and avoided.

The most important features of the implementation model of the programme can be captured, enabling delivery that adheres to the original model and helping to prevent programme drift during maturation and scaling

Source: Gbate, D. (2018), “Developing theories of change for social programmes: co-producing evidence-supported quality improvement”, Palgrave Communications, Vol. 4/1, p. 90, <http://dx.doi.org/10.1057/s41599-018-0139-z>.

The quality and availability of data

The quality and availability of non-evaluation specific data (big data, open data, statistical data, programme monitoring data, etc.) is a primordial factor in how easily a policy can be evaluated. Similarly, the quality of data has an important influence on the rigorousness of the resulting evaluation. In order for data to meet the quality criteria to be used for evaluation, it needs to be accurate, verifiable and documented. Furthermore, policy evaluation and evidence informed policy making (EIPM) can be hindered by the lack of available adequate data and the capacity gaps among government departments and agencies to generate it in a format that can be used. Such challenges include understanding what data and data sets currently exist in ministries and how they can be used for policy analysis. Evaluators and analysts are not necessarily aware of all the data that exists nor do they necessarily have access to administrative data, which may be especially true of external evaluators. Another issue could be that departments do not have comprehensive inventories of all their data holdings and knowledge of their quality. Beyond this, there is a broader data governance challenge that corresponds to the capacity of the public sector to generate the data that is necessary to produce evidence and evaluation, which should also, in theory, be facilitated by the increasing digitalisation of public sector processes.¹

Another challenge relates to the use of individual administrative data. Indeed, data protection legislations can also constitute an obstacle to using individual level data to evaluate policies and programmes in some countries, specifically when carrying out statistical analysis and when merging files, which requires access to single identifiers. Political reticence towards sharing evidence on policy impact and effectiveness may also be another barrier in accessing data.

Box 3.2. Potential sources of data used for policy evaluation

Conducting quality evaluation requires quality data, which may come from various sources:

- **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.

Combining different data sources also has the potential to unlock relevant insights for policy evaluation. Applying big data analysis techniques to public procurement data can contribute to creating stronger, sounder and more relevant evaluations.

Sources: based on Results for America (2017), Government Mechanisms to Advance the Use of Data and Evidence in Policymaking: A Landscape Review

Overall, strategies and policies to combine, link and reuse data, as well as to connect actors and decisions within and outside the public sector, are necessary to enable open data to deliver results (OECD, 2019^[107]). Evidence from the OECD OURData Index suggests that the countries achieving better results are those that clearly assign the responsibility to co-ordinate open data policies.

Some countries have sought to develop EIPM strategies by fostering systematic access to, and use of, administrative data. The US and Japan, for example, have both institutionalised and implemented more systematic structural approaches to facilitate evidence informed policy making. They have done this by mobilising institutional resources, promoting internal champions and exploring the possibility to fully use existing data on a systematic basis through significant governance changes.

The presence of an enabling environment for quality and use

Quality and use are also influenced by a wider enabling environment. The incentives and attitudes of potential users and of evaluators toward conducting evaluations, are influenced by:

- the existence of an enabling environment within the evaluation unit and within the institution as a whole
- the wider environment beyond institutional boundaries and the overall evaluation culture (Vaessen, 2018^[104]).

At the level of evaluation units and individual institutions, quality can benefit from managerial independence – when the evaluation unit can take resource decisions independently – and functional independence – when the evaluation unit can decide on what and how to evaluate (see for a discussion of evaluation independence) (Vaessen, 2018^[104]). Decisions about how to use evidence will also be shaped by the internal dynamics of individual government departments, which includes the organisation culture and internal structures or processes that impact how teams work with each other (Shaxson, 2019^[108]).

The wider environment beyond institutional boundaries also affects the use of evidence (OECD, 2020^[99]). This context can refer to the extent to which ministries are networked with other external organisations, such as knowledge brokers, who can support evidence use (Damschroder et al., 2009^[109]; Greenhalgh et al., 2004^[110]).

Cultural and societal factors may also affect the extent to which evidence gets used in policy-making (OECD, 2020^[99]). For instance, societal attitudes towards policy-making, and what and who should contribute to it, can also affect the use of evidence (Newman, Fisher and Shaxson, 2012^[111]). The erosion in trust in traditional institutions and the digital revolution in communication have eroded the authority of science in some instances. Social media and web-based sources can diffuse opinions very quickly, irrespective of whether they are grounded in scientific evidence (OECD, 2017^[112]). Existing examples of challenges in the communication of science, such as vaccination for example, have led to the recognition that 'more facts' are not enough in addressing these challenges (Sinatra, Kienhues and Hofer, 2014^[113]). Therefore, in order to promote use of evidence, policy makers must also address the societal drivers of resistance to the use of evidence, and recognise the emotional, as well as rational, elements of decision-making.

Besides these external factors, the following sections will focus on the factors that are amenable to policy interventions, including the institutions, strategies and tools developed by governments in order to promote evaluation and use.

Promoting quality through good governance and sound methodology

Understanding quality evaluations

To be credible, a policy evaluation must be technically rigorous, as well as be well governed; that is be independent and appropriate for the decision-making process (Robert Picciotto, 2013^[114]). Therefore, quality evaluations are:

- technically and methodologically sound
- well-governed.

On the first hand, independent processes alone do not guarantee that policy evaluations are of high quality: proper design, sound data collection, rigorous methods, adequate resources are also required. Independent but technically weak evaluations can lead to poor evidence, which can be costly and misleading.

On the other hand, technical quality is necessary but not sufficient to promote an evidence informed approach to policy making. This is because evaluations inherently take place in a political context, as they are usually commissioned by policy and decision-makers, making their outcome susceptible to influence (Parkhurst, 2017^[32]) (Pleger and Hadorn, 2018^[115]). In other words, even when methodologically and technically robust, an evaluation process is 'never truly neutral' (Desautels and Jacob, 2012^[116]). The evaluations of policies and programmes can suffer from a range of biases, whether technical or political, which can affect the evidence-production process. Conversely, evaluations are but one input into policy making and policy and practice decisions must also weigh broader considerations, such as ethics, equity, values and political considerations (Parkhurst, 2017^[32]). The academic literature includes rich discussions of the governance challenges relating to evaluations that may affect the quality of evidence collected ((Barnett and Camfield, 2016^[117]) (Jacob and Boisvert, 2010^[118]) (Brown and Newman, 1992^[119]). Still, the question remains as to what governments can do to promote quality in practice, which will be presented below.

Overview of mechanisms to promote quality evaluations

A large majority of surveyed countries (29 of 42 respondent countries, of which 24 OECD countries) have put in place one or several mechanisms in order to promote quality through various means – thus suggesting that survey respondents have recognised the importance of ensuring the good quality of evaluations.

In general, countries have sought to promote the quality of evaluations via four main determinants:

- developing standards on the quality of the evaluation process, which can be embedded in evaluation guidelines or in legal/policy frameworks
- controlling the quality of the evaluation end product
- supporting and promoting evaluator competences
- fostering quality at an institutional level.

Quality standards for the evaluative process

Firstly, countries have developed mechanisms to ensure that evaluations are properly conducted, that is to say that the *process* of evaluating a policy respects certain quality criteria. In order to do so, countries have developed quality standards, which serve to impose a certain uniformity in the design and process of evaluations (Picciotto, n.d.^[120]).

In many countries, standards for good quality evaluations are embedded in guidelines, which are non-binding documents or recommendations that aim to support governments in the design and implementation of a policy and/or practice (examples include white-books and handbooks). Fewer countries, on the other hand, have embedded such standards in policy or legal frameworks, or normative instruments.

The results of the survey show that most countries have developed standards regarding both the *technical quality* of evaluation and its *good governance*, reflecting their understanding of the dual determinants of quality evaluations. Nevertheless, large differences remain across OECD countries in the content of these guidelines and norms. An analysis of the existing standards for the design, implementation and evaluation of specific public interventions will also complement this analysis (OECD, forthcoming^[106]).

Table 3.1. Mechanisms for the promotion of quality

	Provisions expressed in a policy/legal framework	Guidelines for policy evaluation across government	Competence requirements for evaluators	Peer review (internal/external) of evaluations	Systematic and meta-evaluations	Other
Australia	○	●	○	○	○	●
Austria	●	●	●	●	○	●
Belgium	○	○	○	○	○	○
Canada	●	●	●	●	○	●
Chile	●	○	●	○	○	○
Czech Republic	○	●	○	○	○	○
Denmark	○	○	○	○	○	●
Estonia	●	●	●	○	○	○
Finland	●	●	●	●	○	○
France	●	●	●	●	○	○
Germany	●	●	○	●	●	●
Great Britain	○	●	●	●	●	○
Greece	●	●	●	○	○	○
Hungary	○	○	○	○	○	●
Iceland	○	○	○	○	○	○
Ireland	○	●	○	○	○	●
Israel	○	○	○	○	○	○
Italy	○	●	○	○	○	●
Japan	●	●	●	●	○	○
Korea	●	●	●	○	○	○
Latvia	●	●	○	○	○	○
Lithuania	○	●	○	○	○	○
Mexico	●	●	●	●	●	○
Netherlands	○	●	○	●	○	○
New Zealand	○	●	○	○	○	○
Norway	○	●	○	○	○	○
Poland	●	●	○	○	●	●
Portugal	○	●	○	○	●	○
Slovakia	○	●	●	○	○	○
Slovenia	○	○	○	○	○	○
Spain	○	●	○	●	○	●
Sweden	○	○	○	○	○	○
Switzerland	○	●	○	○	○	○
Turkey	○	○	○	○	○	○
United States	●	●	●	○	○	○
OECD Total						
● Yes	14	26	13	10	5	10
○ No	21	9	22	25	30	25
Argentina	○	●	●	○	○	○
Brazil	○	●	○	○	○	●
Bulgaria	○	○	○	○	○	○
Colombia	●	●	●	○	○	●
Costa Rica	●	●	●	●	●	○
Kazakhstan	○	●	○	○	○	○
Romania	●	○	●	○	○	○

Note: n=42 (35 OECD member countries). 14 countries (12 OECD member countries) answered that there are no mechanisms to ensure the quality of evaluations across government. Answers reflect responses to the questions “How does your government ensure the quality of evaluations across government?” and “Are there guidelines available to support the implementation of policy evaluation across government?”. Systematic and meta-evaluations refer to the evaluations designed to aggregate findings from a series of evaluations. In the option “others”, In Brazil some ministries promote the training of evaluators through its schools of government, and by making available the findings of their evaluations and databases on public sites., In Germany, regular exchange take place within the network of evaluation units of development cooperation agencies and externally through the OECD DAC evalnet. Hungary has a consultation process to review the evaluations, In Ireland each Accounting Officer is responsible for ensuring compliance with the Public Spending Code in their Department/Office. Italy has different mechanisms to improve the quality of the evaluations as part of the National Evaluation system such as steering groups. Poland has a system of assessment of quality of conducted evaluations in the policy evaluation guidelines.

Source: OECD Survey on Policy Evaluation (2018)

Standards set-out in guidelines including provisions for technical quality

A majority of countries (20 countries, of which 17 OECD countries) have developed guidelines that seek to address both the technical quality of evaluations and the good governance of evaluations. Seven countries have developed a single reference guideline for public sector evaluations. Other countries have chosen to adopt distinct guidelines for standards of good governance and for standards regarding methodological rigor. In Estonia, for instance, the Methodology of Impact Assessment (2012) guidelines describe the technical features of impact evaluations of policies and programmes, while the Good Public Engagement Code of Practice (2012) focuses on the principles for the good governance of evaluations, such as the involvement of the public and interests groups in decision-making processes.

International organisations have also adopted such guidelines in order to set standards for quality evaluations and the appropriate principles for their oversight (United Nations Evaluation Group, 2016^[121]). The international organisation that brings together Supreme Audit Institutions has done so as well (INTOSAI, 2010^[122]). At the OECD, the Development Assistance Committee’s *Quality Standards for Development Evaluation* (OECD, 2010^[48]) include overarching considerations regarding evaluation ethics and transparency in the evaluation process, as well as technical guidelines for the design, conduct and follow-up of development evaluations by countries. The OECD Best practices on *ex post* evaluations of regulations (OECD, 2018^[123]) also provide standards relating to the *ex post* evaluation of laws and regulations, and the OECD best practice principles for regulatory policy on Regulatory Impact assessment (GOV/RPC (2018)12/REV2) include provisions for the *ex ante* assessment of regulatory impacts. Similarly, the World Bank Group *Evaluation Principles* sets out core evaluation principles for selecting, conducting and using evaluations (World Bank et al., 2019^[124]) aimed at ensuring that all World Bank Group evaluations are technically robust, as well as credible.

Table 3.2. Quality standards included in evaluation guidelines

	Technical Quality of evaluations						Good Governance of evaluations		None Of The Above
	Identification and design of evaluation approaches	Course of action for commissioning evaluations	Establishment of a calendar for policy evaluation	Identification of human and financial resources	Design of data collection methods	Quality standards of evaluations	Independence of the evaluations	Ethical conduct of evaluations	
Australia	○	○	○	●	○	○	○	○	○
Austria	○	○	○	●	●	○	○	○	○
Canada	●	○	○	○	●	●	●	●	○
Czech Republic	●	○	○	○	●	●	●	●	○
Estonia	●	●	○	●	●	●	●	●	○
Finland	○	●	●	○	○	●	●	●	○
France	●	○	○	○	○	○	○	○	○
Germany	●	●	●	●	●	●	●	●	○
Great Britain	●	○	●	●	●	●	●	●	○
Greece	●	●	●	●	●	●	●	○	○
Ireland	●	○	○	○	●	●	●	○	○
Italy	○	●	○	●	○	○	●	○	○
Japan	●	○	●	○	●	●	○	○	○
Korea	●	○	●	○	●	●	○	○	○
Latvia	●	●	●	●	●	●	○	○	○
Lithuania	●	○	○	●	●	○	●	○	○
Mexico	●	●	●	○	○	●	●	●	○
Netherlands	○	○	○	○	○	●	○	○	○
New Zealand	●	●	○	●	●	●	●	●	○
Norway	●	○	○	●	●	○	○	○	○
Poland	○	○	●	○	●	●	●	○	○
Portugal	○	●	○	○	○	○	○	○	○
Slovakia	●	○	○	○	○	●	●	○	○
Spain	●	●	○	●	○	●	●	●	○
Switzerland	○	○	●	●	●	●	●	●	○
United States	●	○	●	●	●	●	●	●	○
OECD Total									
● Yes	18	10	11	14	17	19	17	11	0
○ No	8	16	15	12	9	7	9	15	26
Argentina	○	●	●	○	●	○	○	○	○
Brazil	●	●	○	●	●	●	●	○	○
Colombia	●	●	○	○	○	○	●	●	○
Costa Rica	●	●	●	●	●	●	●	●	○
Kazakhstan	○	○	●	○	○	○	○	○	○

Note: n=31 (26 OECD member countries). 11 countries (9 OECD member countries) answered that they do not have guidelines to support the implementation of policy evaluation across government. Answers reflect responses to the question, "Do the guidelines contain specific guidance related to the: (Check all that apply)".

Source: OECD Survey on Policy Evaluation (2018)

Identification and design of evaluation approaches

About two thirds of countries (21 countries, of which 18 OECD countries) have included provisions for the design of evaluation approaches in their guidelines. The Spanish State Agency for the Evaluation of Public Policies and Quality of Services (AEVAL), for instance, implemented a practical guide for the design and implementation of public policy evaluation in 2015. The guidelines seek to provide theoretical and practical advice for better evaluation approaches and include detailed recommendations for evaluation design, for instance by proposing key steps for drawing out an intervention's theory of change or illustrating common scenarios for evaluators with local examples.

A further analysis of country guidelines show that these recommend that the purpose, scope (for example time-period, target population, geographic area included, etc.) and objectives of an evaluation be clear. These guidelines underline the importance of making sure that the questions that the evaluation intends to answer are clear and well-defined, as the evaluation criteria and questions define the evidence that the evaluation will generate. Some guidelines also emphasise that the analysis conducted to answer the evaluation question should be clearly and explicitly stated and explained (OECD, forthcoming^[106]).

Box 3.3. The example of the Magenta Book in the United Kingdom: Core questions of policy evaluations

In the United Kingdom, the Magenta book provides guidance on what to consider when designing an evaluation. It invites analysts to consider a series of question such as

- Should it work? (theory of change) What is the underlying 'theory of change', which explains how the policy will make an impact? An understanding of the theory of change that underpins the project will ensure that we measure the things that really matter during the evaluation.
- Can it work? How was the policy delivered (process/implementation evaluation)? How was the policy implemented? Has the policy been properly implemented? What were the challenges to implementation and how were they overcome?
- Does it work? (impact evaluation) Many of our evaluations investigate the impact of the intervention.
- Is it worth it? Do the benefits justify the costs (economic evaluation)? It is anticipated that, if successful, policies/interventions might receive a wider roll-out. It will therefore be important to consider whether they are cost effective.

Source: (HM Treasury, 2011^[9]). www.gov.uk/government/publications/the-magenta-book

Finally, an evaluation plan or matrix may be a useful tool to lay out the evaluation's focus, the main questions it seeks to answer, the key information needed for indicators, data collection methods, etc. Importantly, such an evaluation plan should also mention the purpose of the evaluation and how its results should be put to use (Kusters et al., 2011^[125]). The Lithuanian ministry of finance, for example, issued *Recommendations on Implementation of Programs Evaluation* (2011), which give advice on how to plan and design an evaluation, from identifying the need for an evaluation to establishing an evaluation plan, including methods.

Course of action for commissioning evaluations

Some country guidelines (14 countries overall, of which 10 OECD countries) include specific standards or recommendations regarding the commissioning of evaluations, as is the case in Costa Rica, where the ministry of national planning and economic policy (Mideplan) has dedicated a separate guideline for the establishment of an evaluation's terms of reference.

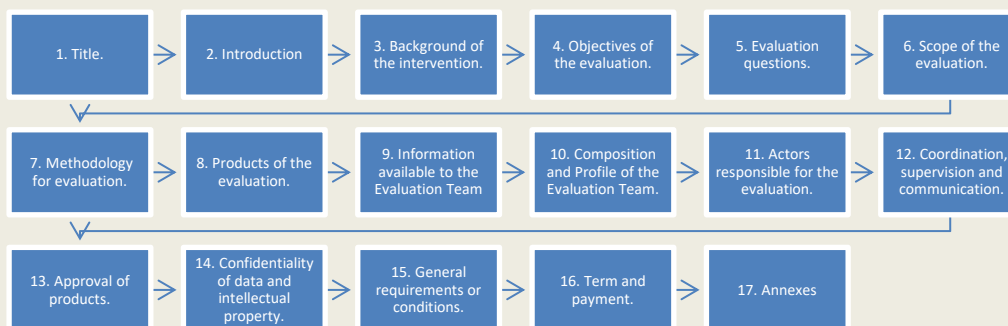
Box 3.4. Standards for commissioning evaluations in Costa Rica

The ministry of national planning and economic policy (Mideplan) has developed specific guidelines and standards for the preparation of Terms of Reference for policy evaluations (*“Guía de Términos de Referencia: Orientaciones para su elaboración: estructura y contenido”*). This handbook provides inputs on the recommended technical content and basic structure of terms of reference (ToRs) for policy evaluation that are commissioned to an external agent. The methodological tool is composed of two main parts:

5. what the ToRs are and what they are for.

the basic structure that this document should present and the essential characteristics of its content.

According to the document, terms of reference should include at least the following criteria:



Finally, the guidelines recommend that the ToR be clear, concrete and that the main actors involved in the public intervention and evaluation submit them for consultation and validation.

Source: (Mideplan, 2018^[126]).

In fact, drafting the terms of reference (ToRs) provides the guidelines for the work that will have to be carried-out during the evaluation process and therefore constitute an essential tool for quality assurance (Kusters et al., 2011^[125]). The terms of Reference is an essential document of any evaluation. Country guidelines mention that ToRs should likely cover the background context of the evaluation, its scope, goals, methodology, team composition, stakeholders to be engaged and the evaluation budget (Independent Evaluation Office of UNDP, 2019^[127]). Evaluation guidelines developed by countries may also specify that ToRs should be drafted by the evaluation manager once the relevant data and documents are collected.

Planning out evaluations and identifying the appropriate resources

Good evaluation planning may also be important to ensure quality, as well as use. Many researchers emphasise the importance of the issue of timeliness of evaluation results to promote their use in decision-making (Leviton and Hughes, 1981^[100]): the consensus is that evaluations should be thought of well in advance and the evaluation process planned-out carefully. Likewise, resource limitations can

strongly influence an evaluation's impact and use, making the identification of human and financial resources an important step in planning-out the evaluation process.

However, only a minority of countries (14 countries, of which 11 OECD countries) include clauses regarding the establishment of a calendar for policy evaluation in their guidelines. Similarly, less than half of OECD countries and of overall respondents (16 countries, of which 14 OECD countries) include standards regarding the identification of human and financial resources for evaluation in their guidelines.

One notable exception is Korea's office for government policy coordination (2017) framework act on government performance evaluation, which recommends a systematic approach to evaluation on a yearly basis in order to facilitate planning out resources for this purpose.

Box 3.5. Korea's office for government policy Coordination (2017) framework act on government performance evaluation

The framework act on government performance evaluation recommends that all government agencies formulate a **yearly internal evaluation plan** to identify the major policies to undergo review each year. The results of the evaluations are to be submitted to the government performance evaluation committee (GPEC) in the spring. Such evaluation plans may allow evaluators to adequately plan the necessary resources for the evaluations, as well as to ensure that the results of the evaluations are useful to decision-makers, as the timeline for the publication of their results is clear.

The document gives **specific instructions regarding the composition of the GPEC**, which is in charge of implementing government performance evaluation:

- It should be composed of no more than fifteen members including two chairpersons.
- Members should have earned a degree in a discipline related to evaluation and have experience related to academia or a research institute.
- Members who are not government officials should be into office for a two years term, and may only serve one additional consecutive one.

Source: Korea Office for Government Policy Coordination (2017), *Framework Act on Government Performance Evaluation*.

Design of data collection methods

A majority of countries (20 countries, 17 OECD countries) have included standards for the design of data collection methods in their guidelines. Indeed, although experimental approaches, such as randomised control trials, can sometimes take advantage of existing administrative data, it is often necessary to collect new data for an evaluation, using social research methods.

The design of data collection methods is key for conducting policy evaluations and many country guidelines include data collection standards: they should be representative, their approach should be well-designed, and their size should be appropriate (OECD, forthcoming^[106]). Moreover, guidelines prescribe that data should be opened to critical questioning and challenge, and its use to conduct an analysis should be explained and justified. This analysis should be subsequently well-defined and executed (see examples for France and Norway in (Box 3.6).

Box 3.6. Data collection standards in guidelines

French Guidelines on evaluating the impact of public policies

The French guidelines for decision makers and practitioners on how to evaluate the impact of public policies (*Comment évaluer l'impact des politiques publiques: un guide à l'usage des décideurs et des praticiens*, 2016) underline how the quality of an impact evaluation depends on the availability, breadth and quality of data on the policy being evaluated. According to the guidelines, creating relevant indicators to measure the impact of a policy requires access to various data sources and variables, and thereby a frequent matching of statistical sources. The right type of data should be collected for a valid implementation of the evaluation method chosen. For example, the guidelines describe how, when using the matching method to establish the causal effect of a policy on certain outcomes, data on individuals and their social and economic environment has to be sufficiently rich to minimise selection bias.

The guidelines recommend that, when using qualitative data (from surveys, field observations or case studies), the credibility of results be increased by comparing and combining information from different actors and methods. These guidelines conclude on the need to institutionalise and better operationalise the production of and access to data. Examples of processes to promote data access include the accelerated provision of administrative files and the facilitated access procedures to institutions such as the National Council for Statistical Information (CNIS). Lastly, the guidelines note the virtue of conducting a systematic review of the readily available to assess whether collecting new or existing data is needed in the first place. In addition, France has created a secure access to statistical and administrative micro data through a single entry point to a large number of data producers (www.casd.eu/en/).

Norwegian guidelines on carrying out evaluations

These guidelines give an explicit methodology for collecting data, from choosing the collection instrument (survey, interview, observation, etc.) and the subjects (individuals, businesses, etc.) to obtaining the information, and registering and processing the data. They give advice regarding the choice of data collection and analysis methods, and recommend to combine them to increase the quality of a single evaluation. Among other examples, a precise step-by-step guide is provided on how to conduct a survey, one of the most common data collection methods. This guide includes suggestions on reaching out to as many relevant respondents as possible, designing clear and precise questions and achieving a high response rate. Finally, these guidelines emphasise the importance of evaluators' analytical knowledge and skills to ensure the correct use of data and avoidance of data saturation.

Sources: (France Stratégie, 2016^[128]), Norway Ministry of Finance (2005), *Guidelines on carrying out evaluations*

Evaluation methods

Choosing the appropriate evaluation method is paramount to an evaluation's quality. A high quality evaluation method solves the issue of attribution (*causality*) by providing insights on whether and to what extent a policy delivered its intended outcomes.

Table 3.3. Impact evaluation methods

	Evaluation method	Description	Limits
Quasi experimental	Pre-Post	Impact is measured as the change in the outcomes of participants before and after the policy is implemented.	Factors other than the policy itself that might have influenced the outcomes of participants are not accounted for.
	Simple Difference	Outcomes of participants and non-participants after the policy is implemented are compared.	Results are biased if participants and non-participants have different chances of being affected by the policy before its implementation, and if they differ in other ways than their participation status.
	Differences in Differences	The policy effect is measured by comparing the evolution of the participants' outcomes before and after its implementation with the evolution of non-participants' outcomes throughout that same period.	There will be bias if the control group does not actually reflect what would have happened to the treatment group had it not been treated. For valid results, the observable and unobservable differences between the two groups should also be constant across time.
	Multiple Linear Regression	This method consists of comparing the outcomes of participants and non-participants, controlling for observable differences between the two groups that might affect their outcomes (gender, income, education, age, etc.).	Unobservable, unmeasurable and unmeasured factors may still differ across the two groups and affect the measured outcome, which would limit the validity in estimating the causal impact of the programme.
	Statistical Matching	Participants and non-participants who have otherwise similar characteristics are compared.	Unobserved, unmeasurable and unmeasured characteristics may still bias the estimated effect.
	Regression Discontinuity Design	Individuals are ranked according to a given measurable criteria, and a cut-off determines their participation in the policy. Participant just above the cut-off are compared to non-participants just below.	There is a risk that individuals manipulate their own outcomes to become eligible (or not) for the policy, which introduces bias. Moreover, the measured effect is only "local", meaning that it holds only for individuals close to the cut-off.
	Instrumental Variables	The effect is measured by identifying an "instrumental" variable that affects the outcome of interest only indirectly through determining whether an individual participates in the policy. This instrument should not be related to any other factor affecting the outcome of interest.	The validity of results relies on finding a good instrument, or one that predicts the outcome only through programme participation, which is difficult in practice.
Statistical	Randomised Evaluation	This experimental method consists in randomly assigning individuals to participate in the policy or not, and comparing outcomes of the two groups. Random assignment removes, on average, any differences between the participants and the non-participants, apart from their participation status.	Causal estimation from randomised evaluation is valid if only if randomisation was "properly" conducted. Examples of bias are that the effect on the treatment group "spilled over" on the control group (spill over effects), or that treated individuals ended up not participating in the programme (attrition bias).

Source: Source: Based on J-Pal (2016) *Impact Evaluation Methods: What are they and what assumptions must hold for each to be valid?*

Specifically, impact evaluation methods provide a solid counterfactual, that is to say take into account all the other factors that could generate an observed outcome (Campbell and Harper, 2012^[129]). The question of what approach is most appropriate will depend on the complexity of the relationships between an intervention's inputs, activities, outputs, outcomes and impacts.

According to OECD data, two thirds of countries (21 countries overall, of which 19 OECD countries) include provisions detailing quality standards for evaluation methods (see column "Quality standard for evaluations"). Some countries have developed methodological guidebooks or manuals with the primary intent of delving deeper into evaluative methods in order to provide evaluators with practical advice for the implementation of an evaluation. In Great Britain, for instance, several methodological handbooks provide detailed recommendations on how to evaluate policy impacts and conduct programme appraisals and evaluations. Other countries that have developed such handbooks or detailed guidelines on evaluative methods include France (France Stratégie, Desplatz and Ferracci, 2016^[130]), Spain (AEVAL, 2015^[131]) and Lithuania (Ministry of Finance (Lithuania), 2011^[132]).

Box 3.7. Evaluation guidelines in Great Britain

The UK Government has been committed to improving central and local government efficiency and effectiveness through the development of different tools to ensure public policies are based on reliable and robust evidence. To achieve this, the HM Treasury's Green and Magenta Books together provide detailed guidelines, aimed at policy makers and analysts, on how policies and projects should be assessed and reviewed, which makes the two sets of guidance complementary.

The Magenta book: guidance for evaluation

The Magenta Book comprises central government guidance on public policy evaluations. It presents standards of good practice in conducting evaluations, and seeks to provide an understanding of the issues faced when undertaking evaluations of projects, policies, programmes and the delivery of services.

The green book: central government guidance on appraisal and evaluation

The Green Book is guidance issued by HM Treasury on how to *appraise* policies, programmes and projects. It also provides guidance on the design and use of monitoring and evaluation before, during and after implementation. A range of templates and guidance on specific analysis topics and analysis techniques, which are frequently encountered during government analysis, are found in the Aqua book.

The Aqua Book: guidance on producing quality analysis for government

The Aqua Book is a suite of resources aimed at improving analytical quality assurance. Combining the high-level principles of analytical quality assurance, together with clarified roles and responsibilities, the Aqua Book helps departments and agencies embed an analytical environment that assists the delivery of quality analysis, deliver greater consistency in the approach to analytical quality assurance processes across government and ensure that the commissioners of analysis have greater confidence in analysis.

In practice, nevertheless, only a minority of countries use impact evaluation methods, such as randomised controlled trials, to evaluate their government-wide policy priorities (8 countries, of which 7 OECD countries).

Table 3.4. Methods used by countries in the evaluation of government-wide policy priorities

	Regression/econometrics/structural equation modelling	Randomised controlled trials	Qualitative Comparative Analysis	Contribution analysis	(Comparative) case studies	Process tracing	Theory-based evaluation
Australia	○	○	○	○	○	○	○
Austria	○	○	○	○	○	●	○
Canada	●	○	●	●	●	○	●
Chile	○	○	●	○	●	○	○
Estonia	●	○	●	●	○	○	○
Finland	○	○	●	●	●	●	○
France	●	●	●	○	●	○	○
Germany	●	●	●	●	●	●	●
Great Britain	●	●	●	●	●	●	●
Greece	●	○	●	●	●	●	○
Hungary	○	○	●	○	○	●	○
Ireland	●	●	●	○	●	○	○
Israel	○	○	○	○	○	●	○
Italy	○	●	○	○	●	○	○
Japan	○	○	○	○	●	○	●
Korea	○	○	○	○	○	○	○
Lithuania	●○	○	●	○	●	○	●○
Latvia	●	○	●	○	○	○	●
Mexico	●	●	●	●	○	●	●
Poland	●	●	●	●	●	●	●
Portugal	○	○	○	○	○	●	○
Slovakia	●	○	○	○	○	○	●
Spain	●	○	●	○	○	●	○
Sweden	●	○	○	○	●	○	●
OECD Total							
● Yes	13	7	15	8	13	11	9
○ No	11	17	9	16	11	13	15
Argentina	○	○	○	○	○	○	○
Brazil	●	○	●	○	●	●	●
Colombia	●	●	●	○	●	○	○
Costa Rica	●	○	●	●	●	●	●
Romania	●	○	○	○	○	○	○

Note: For the main institution on government-wide policy priorities n=29. 4 countries answered that they do not have government-wide policy priorities. Moreover, 9 countries answered that they do not evaluate their government-wide policy priorities. Answers reflect responses to the question, "Which quantitative or qualitative methods of impact evaluation have been used over the past three years for the evaluation of government wide policy priorities? (Check all that apply)".

Source: OECD Survey on Policy Evaluation (2018)

Guidelines for the good governance of the evaluation process

Individuals and organisations conducting policy evaluations also need to ensure the credibility of the evidence produced by putting in place mechanisms to promote the integrity of the evaluation process (OECD, forthcoming^[106]). In fact, an evaluation's impact can depend on its perceived quality, in terms of its readability and perception of transparency and lack of bias, as much as it can on its technical quality. Stakeholders and an evaluation's clients must therefore trust its findings and find them credible (Caroline Heider, 2018^[133]).

Independence of evaluations

Firstly, the independence of process for conducting policy evaluations is also a crucial element of their credibility (France Stratégie, Desplatz and Ferracci, n.d.^[134]). The notion of independence can be understood as an evaluation being free from undue political pressure and organisational influence. The literature distinguishes between several types of independence: structural, functional and behavioural independence (Vaessen, 2018^[104]) (Robert Picciotto, 2013^[114]).

Box 3.8. Understanding independence in evaluations

Independence in evaluations is a critical element of their credibility and ultimately quality. It consists in evaluations being free and protected from undue political and managerial influence. Three types of such independence are mentioned in literature (Vaessen, 2018^[104]):

- **Structural and functional independence** refer to the independence of the evaluation team with respect to management, both in terms of the object and processes of the evaluation and in the decisions concerning human and financial resources.
- **Behavioural independence** relates to the unbiasedness and integrity of the evaluator.

As such, independence requires avoiding conflicts of interests, complying with ethical norms of conduct and the independence of the evaluation commissioners themselves. In practice, independence is usually difficult to achieve in internal evaluations, where political influence is often exerted and various political interests are at stake. Accordingly, appointing an external evaluator is a common solution to foster more impartial and trustworthy results, but it may not always solve the issue of the pressures from private interests and lobbying efforts, which can implicitly weigh on external evaluators.

Independence can only be pursued to a certain extent, as there is a complex trade-off between evaluation independence and quality. External evaluators are indeed more prone to be free from political biases, but they risk lacking sufficiently thorough and adequate knowledge about the policy being evaluated. They can also be subject to influence by specific private interest groups and may have more difficult access to relevant administrative data. Conversely, internal evaluators have the potential to offer constructive views and expertise thanks to their familiarity with the policy subject and knowledge of its political relevance.

Lastly, managerial influence can also provide effective incentives and positive support so that the results of evaluations are used and understood.

Source: (Picciotto, 2013^[56]), (Vaessen, 2018^[104]), (France Stratégie, 2016^[128]), (Wildavsky, 1979^[135]).

OECD data shows that 20 countries include provisions regarding the independence of evaluations in their evaluation guidelines, understood broadly. While evaluation guidelines usually put emphasis on behavioural independence (i.e. how the evaluator should act to maintain independence in the evaluative process), countries have also put in place other safeguard mechanisms to ensure the structural and functional independence of government evaluators. Refer also to Chapter 2. on the institutionalisation of policy evaluation for a detailed discussion of the subject.

Box 3.9. Australia's productivity commission: An autonomous government body

The Australian government's productivity commission is an autonomous research and advisory body that focuses on a number of economic, social and environmental issues affecting the wellbeing of Australians. At the request of the Australian Government, it provides independent and quality advice and information on key policy and regulatory issues. It also conducts self-initiated research to support the Government in its performance reporting and annual reporting, and acts as a secretariat under the council of Australian government for the inter-governmental review of government service provision.

The commission is located in the Government's treasury portfolio and its activities range across all levels of governments. It does not have executive power and does not administer government programmes. The Commission is nevertheless effective in informing policy formulation and the public debate thanks to three characteristics:

- Independence: it operates under its own legislation, and its independence is formalised through the productivity commission act. Moreover, it has its own budget allocation and permanent staff working at arm's length from government agencies. Even if the commission's work programme is largely defined by the government, its results and advice are always derived from its own analyses.
- Transparent processes: all advice, information and analysis produced and provided to government is subject to public scrutiny through consultative forums and release of preliminary findings and draft reports.
- Community-wide perspective: under its statutory guidelines, the Commission is required to take a view that encompasses the interests of the entire Australian community rather than particular ones.

Source: Australian Government. "About the Commission" and "How we operate". Accessed September 2nd 2019. <https://www.pc.gov.au/about>, <https://www.pc.gov.au/about/operate>

Ethical conduct of evaluators

Standards for ethical conduct of evaluators are found in approximately a third of the sample: out of the 42 countries who responded to the survey, 13 have developed such standards (11 OECD countries). These standards can include provisions for the use of administrative and big data, for instance when issues of consent are raised where information provided by citizens is being used. Other approaches focus on ensuring that evaluators conduct their research and data collection in ways that ensure the safeguard of the dignity, rights, safety and privacy of participants (e.g. OMB guidance).

Finally, standards for the ethical conduct of evaluators include mechanisms focused on the prevention of conflicts of interests. In fact, a key part of standards of public life is that officials do not act or take decisions in such a way as to gain financial or other material benefits. Such principles of ethical conduct are outlined in the US office of management and budget's monitoring and evaluation Guidelines for agencies that administer foreign assistance, which advises the full disclosure of any conflict of interest among evaluators (Office of Management and Budget, 2018^[136]), as well as in the recent programme evaluation standards and practices issued as part of the implementation of the 2018 Act on Evidence Based Policy Making (Table 2.4). Similarly, the Swiss guide for evaluation of the confederation's efficacy also underlines the importance of determining relevant actors' and stakeholders' needs and interests early enough to allow sufficient time to identify and solve conflicts of interest (Office fédéral de la justice, 2005^[137]).

The OECD has also developed ‘Guidelines for managing conflict of interest in the public service’ whose primary aim is to help countries, at the central government level, consider conflict of interest policies and practices. These guidelines pertain to all public officials, in any capacity, and are not necessarily geared towards the evaluators or the producers of evidence.

Box 3.10. Sources of conflict of interest in evaluations

Conflicts of interest often arise when evaluators have **previous or intended future work experience related to the policy being evaluated** (Independent Evaluation Office of UNDP, 2019^[127]). To minimise them, evaluation commissioners may avoid employing evaluators who had prior engagement in the decision-making, financing or design of the policy being evaluated. Evaluators of a particular policy should not be subsequently involved in any service related to that same policy, from implementation to design.

Conflicts of interests may also come from particular **personal relationships** between evaluators and commissioners, such as close family members who may be in a position to influence the evaluation or its outcome on the policy (Picciotto, 2013^[56]). Research has shown that, although often unnoticed, **evaluation clients** can exert pressure on evaluators, which is a source of conflict that may be avoided by improving communication between the two parties (Pleger and Hadorn, 2018^[115]).

Lastly, **lobbyists and advocacy groups** can exert influence to further their particular interest, often at the expense of the public interest. It is nevertheless important to note that these groups also have the capacity to bring valuable information into the evaluation and its related policy debate. Overall, evaluators should follow the principle of full disclosure of any actual or potential conflicts of interest, and procedures should be put in place to identify relationships that might put the objectivity of the evaluation at risk. An example of such procedure is analysing the resumes of current and potential evaluators and circulating them to partners and stakeholders to decide whether they should be dismissed or employed.

Sources: OECD (2019) Meeting of the Coalition of Influencers on Integrity in Public Decision-Making, (Independent Evaluation Office of UNDP, 2019^[127]), (Pleger and Hadorn, 2018^[138]), (Picciotto, 2013^[56]).

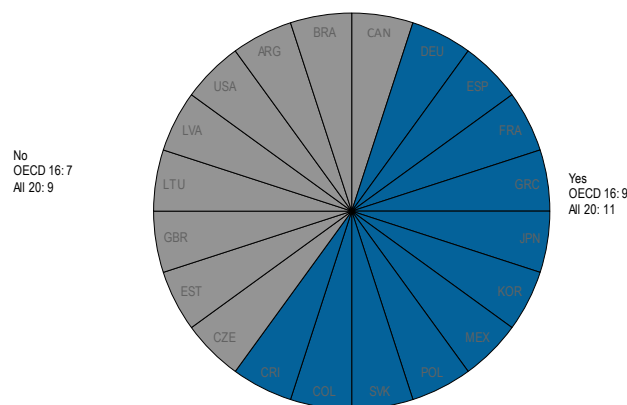
Other standards relating to the good governance of the evaluation process

In addition to the previously mentioned standards for the oversight of evaluations, other standards have been identified as relevant by literature. These include the principles of transparency, accountability, appropriateness and integrity. The OECD is currently conducting a mapping of principles and standards for the good governance of evidence. This exercise required an extensive stocktaking of country and academic experiences to identify a list of core principles for the governance of evidence. These principles, which are equally applicable to the governance of policy evaluations, mainly address issues such as the appropriateness of the evidence, the accountability and transparency of evidence, and the need for evidence to be ready for critical questioning and public scrutiny (OECD, forthcoming^[106]).

Standards embedded in legal frameworks

Some countries have also embedded such standards in their policy or legal framework, meaning that the standards are included in normative instruments. Overall, fewer countries have chosen to embed standards for good quality evaluations in a normative instrument, suggesting that countries view these standards as *recommendations* to be used in a proportional manner by evaluations and managers depending on the local context – rather than fixed rules. For instance, only nine OECD countries have adopted standards for quality methods in their policy/legal framework related to policy evaluations.

Figure 3.1. Countries that have standards for quality evaluations in their policy/legal framework



Note: n=20 (all countries surveyed who have a policy framework, among which 16 are OECD). Answers reflect responses to the question, "Which elements do(es) the document/s referred to under Q4 and Q5 cover concerning policy evaluation across government? (Check all that apply)".

Source: OECD Survey on Policy Evaluation (2018).

The Korea's office for government policy coordination (2017) framework act on Government performance evaluation contains quality standards relating to planning and carrying out evaluations. Likewise, the national evaluation policy (PNE) in Costa Rica seeks to ensure the quality of evaluations by promoting the evaluability of government programmes, increasing the involvement of stakeholders in the evaluative process and establishing competency requirements for evaluators. The Czech Republic, Germany, Spain, Estonia, Great-Britain, Korea, Lithuania, Poland and Costa Rica have also embedded standards related to the ethical conduct of evaluators in their legal framework.

While this report focuses mainly on public sector standards related to the quality of evaluation, there are also many standards established and proposed by the private sector (OECD, forthcoming^[106]).

Measures to control the quality of the evaluation product

In various countries, quality control mechanisms are developed in addition to the standards and guidelines in place to ensure the quality of policy evaluations. Mechanisms for quality control ensure that the evaluation design, as well as its planning and delivery, have been properly conducted to meet the pre-determined quality criteria. While quality assurance mechanisms seek to ensure credibility in how the evaluation is conducted (the process). Quality control tools ensure that the end product of the evaluation (the report) meets a certain standard for quality. Both are key elements to ensure the robustness of policy evaluations (HM Treasury, 2011^[9]). Overall, quality control mechanisms are much less common than quality assurance mechanisms, with only approximately one third of countries (31% of countries overall) using a quality control mechanism (for example, a peer review of evaluations or meta-evaluations). An example of a country with quality assurance mechanisms is Japan; there, the Ministry of Internal Affairs and Communications (MIC) is in charge of quality assurance of policy evaluation. It checks the ministries' evaluations, holds inter-ministerial liaison meetings, uses academic and practical experts' insights, and publicises information about policy evaluation (see Box 3.11).

Box 3.11. The review function of Japanese Ministry of Internal Affairs and Communications (MIC)

The MIC conducts coherent and comprehensive quality controls of policy evaluations done by the ministries. The Administrative Evaluation Bureau (AEB) reviews the evaluations carried out by the ministries, identifying elements that need to be improved and publicised on the basis of the basic guidelines for implementing policy evaluation (Cabinet Decision in 2005 and latest revised in 2017). This includes:

- an examination of the objectivity and rigor of policy evaluations conducted by ministries.
- a determination of the need for the implementation of a new evaluation or further evaluation..
- ensuring objective and rigorous implementation is deemed impossible if left to the ministry.

The role of liaison meetings

The MIC hosts inter-ministerial liaison meetings to foster close communication, ensure the implementation of evaluations and promote initiatives related to policy evaluation, with a view to improving quality.

The use of academic experts

The use of insights of academic and practical experts is aimed at ensuring the objective and rigorous implementation of policy evaluation, thereby assuring quality. Experts' insights are collected through interviews, in various steps including what the policy management cycle such as PDCA by the policy evaluation should be, setting primary goals of policies, and summarising policy evaluation results.

The policy evaluation council

The policy evaluation council established under the MIC investigates and discusses important matters relating to policy evaluation and the AEB investigation. The council is composed of members who have been selected based on their expertise in the academic, administrative and private fields. In regards to policy evaluation, the council discusses important matters relating to the development and revision of guidelines, and the objectivity and rigor of evaluation results.

Publicising information about policy evaluation

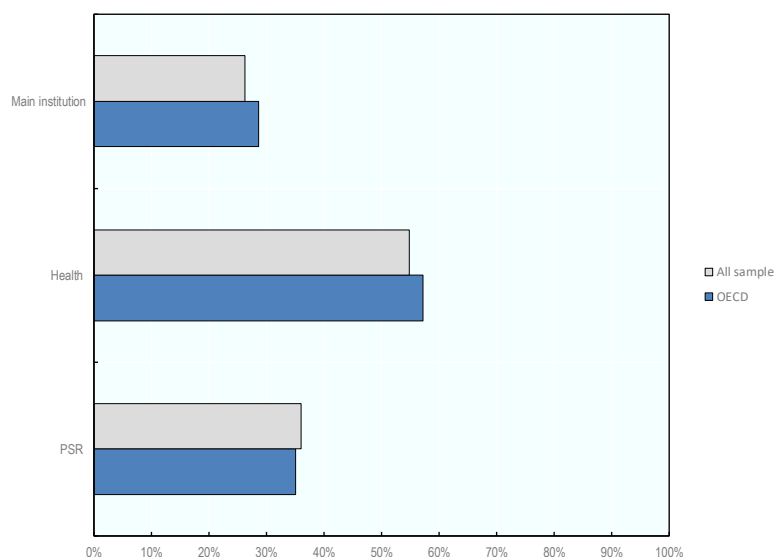
Policy evaluation reports are made public, together with information on how the results are used for the development of policy. The MIC must also publicise its evaluation plan and evaluation reports. The MIC also prepares an annual report on the status of policy evaluation conducted by the ministries and how the results of the evaluation have been reflected in policymaking process, which must be publicised and reported to the National Diet of Japan. The MIC aggregates results of policy evaluation by the ministries on the Portal Site for Policy Evaluation. This system leads to contribution to ensuring the quality of evaluation as well as accountability and transparency of implementation.

Source: Ministry of Internal Affairs and Communications (Japan).

Peer review of evaluation products

The most common control mechanism used by countries to promote quality of evaluations is the peer review process. Peer reviews consist of a panel or reference group, composed of external or internal experts, subject an evaluation to review of its technical quality and substantive content. The peer review process helps determine whether the evaluation meets the adequate quality standards and can therefore be published, as illustrated by examples for Portugal and Germany.

Figure 3.2. Peer reviews



Note: For the main institution n=42 (35 OECD member countries). For the Health ministries n=31 (28 OECD member countries). 9 countries (7 OECD member countries) did not participate on this survey. Moreover, 2 countries (1 OECD member country) are not included as they answered that none of the policies that fall in their institution's responsibility are evaluated. For the Public Sector Reform (PSR) ministries n=25 (20 OECD member countries). 11 countries (10 OECD member countries) did not participate on this survey. Moreover, 6 countries (5 OECD member countries) are not included as they answered that none of the policies that fall in their institution's responsibility are evaluated. Answers reflect responses to the question "How does your government ensure the quality of evaluations across government".

Source: OECD Survey on Policy Evaluation (2018).

Box 3.12. Internal and external peer reviews in Portugal and Germany

Evaluation of the Portuguese Simplex Programme

The evaluation of the Portuguese Simplex programme shows a form of combined (internal and external) peer reviews. On the internal level, project managers are required to regularly report on the progress of the project plan. Reporting is done through an electronic platform and during meetings with key stakeholders and partners, allowing for relevant internal and external insights. The results of such reviews are uploaded on a publicly accessible website, so that citizens may also have a critical say on the advancement of the programme and on the progress report shared every trimester. At the same time, external contractors such as academics from Nova University or evaluators from the European Commission evaluate the programme.

Evaluation of the German Strategy on Sustainable Development Goals

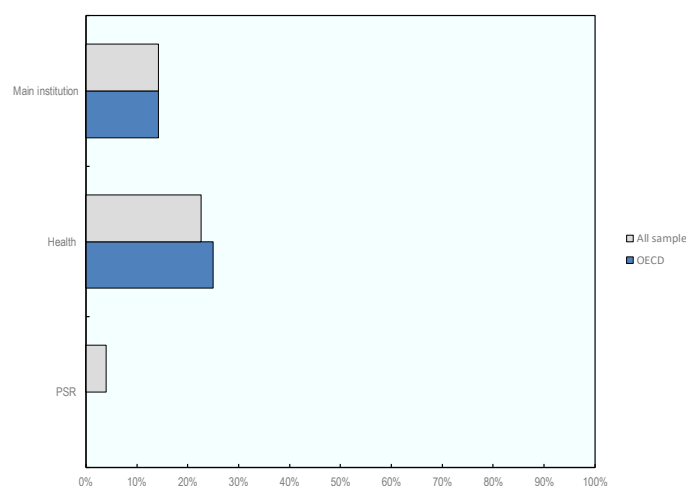
The German Chancellor invited an international group of recognised experts to review the country's 2013 Sustainability Strategy. Following a first external peer review made in 2009, this one includes a variety of experts such as Korean and German experts, the former senior vice President of Unilever, members of Parliaments, Chair of WWF South Africa, or again the former director general of the UK Department of Environment. Such diverse group of peer review may provide constructive insights on the evaluation of the German Strategy, allowing for further improvements in its quality, and ultimately its quality.

Source: OECD (2018) Survey on Policy Evaluation, German Council for Sustainable Development (2013) *Peer review: Germany leads the way* <https://www.bundesregierung.de/breg-en/issues/sustainability/peer-review-germany-leads-the-way-402952> (Accessed 28th of August 2019).

Meta-evaluations

Meta-evaluations correspond originally to the evaluation of an evaluation to control its quality and/or assess the overall performance of the evaluation (Scriven, 1969^[139]). Nowadays, it mainly refers to evaluations designed to aggregate findings from a series of evaluations. In its latter meaning, meta-evaluation is an evidence synthesis method (see section on ‘Methods for Reviewing and Assessing the Evidence Base’ for other evidence synthesis methods), which serves to evaluate the quality of a series of evaluations (by making an assessment of evaluations through reports and other relevant sources) and its adherence to established standards. As such, meta-evaluations constitute a useful tool to review the quality of policy evaluations before they are made publicly available. The figure below shows that a relatively limited number of countries use meta-evaluations to control the quality of evaluations. This might either be due to a lack of skills, familiarity or methods.

Figure 3.3. Systematic and meta-evaluations



Note: For the main institution n=42 (35 OECD member countries). For the Health ministries n=31 (28 OECD member countries). 9 countries (7 OECD member countries) did not participate on this survey. Moreover, 2 countries (1 OECD member country) are not included as they answered that none of the policies that fall in their institution's responsibility are evaluated. For the PSR ministries n=25 (20 OECD member countries). 11 countries (10 OECD member countries) did not participate on this survey. Moreover, 6 countries (5 OECD member countries) are not included as they answered that none of the policies that fall in their institution's responsibility are evaluated. Answers reflect responses to the question “How does your government ensure the quality of evaluations across government”. Systematic and meta-evaluations refer to the evaluations designed to aggregate findings from a series of evaluations. It can also be used to denote the evaluation of an evaluation to judge its quality and/or assess the performance of the evaluation. Countries that reported no mechanisms to ensure the quality of evaluations across government are equal to zero.

Source: OECD Survey on Policy Evaluation (2018).

Box 3.13. Meta-evaluations

A **meta-evaluation** is a systematic, managed and controlled method to assess the quality of processes and results of carried out evaluations (Malčík and Seberová, 2010^[140]). Meta-evaluations can take several forms:

- Formative meta-evaluations intend to guide a primary evaluation (Stufflebeam, 1978^[141]). In this dimension, the meta-evaluation is used as an instrument to improve or change an ongoing evaluation design and implementation (Better evaluation, 2019^[142]).
- Summative meta-evaluations denote studies that judge the merits of completed evaluations (Better evaluation, 2019^[142]). This dimension is connected to ensuring the quality, validity and correctness of the primary evaluation, thus verifying whether key principles have been followed and whether its results can themselves be judged as relevant, valid and reliable.

Source: in the text

An exception is the meta-evaluation in Costa Rica, led by the program for the promotion of capacities in evaluation in Latin America (*Programa de Fomento de Capacidades en Evaluación en diversos países de América Latina* - FOCEVAL) in 2016. This meta-evaluation sought to assess the usefulness of a set of evaluations², their methodological rigor, their success in resource management, and their professional and ethical performance. The meta-evaluation provided relevant information to the ministry of national planning and economic policy (Mideplan) to improve stages in the evaluation process, such as an agreement among institutional authorities to reduce the times for starting the evaluation, and an enhancement of the terms of reference to promote a more rigorous and clear evaluation process.

Self-evaluation tools and checklists

Finally, some countries have also developed tools aimed either at the evaluators themselves (i.e. self-evaluation) or at the managing and/or commissioning team (quality control checklists, for example) in order to help them control whether their work meets the appropriate quality criteria.

Quality control checklists are aimed at standardising quality control practices of evaluation deliverables and as such can be useful to evaluation managers, commissioners, decision-makers or other stakeholders to review evaluations against a set a pre-determined criteria (Stufflebeam, 2001^[143]). The evaluation unit in the European Commission, for example, includes a clear quality criteria grid in its terms of reference, against which the evaluation manager assesses the work of the external evaluators (OECD, 2016^[144]).

Self-evaluation, on the other hand, is a critical review of project/programme performance by the operations team in charge of the intervention, as they serve to standardise practices when reviewing evaluation deliverables. Although less commonly used (only two respondent countries mentioned their use), self-evaluation tools can form an important element of a quality control system (OECD, 2016^[144]), as they constitute the first step in the control process.

Box 3.14. Self-evaluation checklists in Spain and Poland

Only two countries reported the use of a self evaluation checklist, the results of which are presented below:

The Spanish institute for the evaluation of public policies' auto-satisfaction survey

The Spanish institute for the evaluation of public policies (IEPP, formerly AEVAL) has developed an auto-satisfaction survey, whereby participants in the evaluation share their satisfaction regarding the evaluation process and quality. This stage of the evaluation follow-up process favors responsiveness to the evaluation client by providing specific measurements of the quality and degree of usefulness of evaluation products such as the evaluation report.

The Polish Ministry of Infrastructure and Development's self-assessment checklist

This self-assessment checklist, presented in the national Guidelines on evaluation of cohesion policy for 2014-2020, aims to prevent implementing recommendations from poor quality evaluations. This system is one of the components of meta-evaluations, focusing on the skills and practices of the evaluators rather than the evaluation more broadly. The checklist includes criteria such as the extent to which the objectives were achieved, the methodology used and the data reliability. Each criteria is given a numerical rating that can be supplemented with qualitative comments (Polish Ministry of Infrastructure and Development, 2015^[145]).

Sources: OECD (2018) Survey on Policy Evaluation, Polish Ministry of Infrastructure and Development (2015), Self-Assessment Checklist.

Promoting competencies for policy evaluation

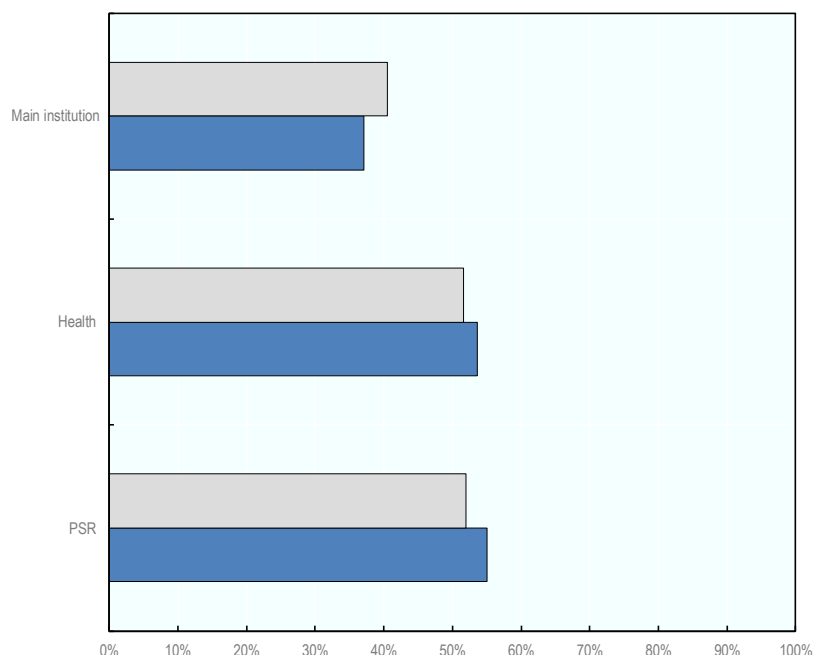
While quality guidelines and standards provide evaluators with resources to help them make the appropriate decisions when conducting evaluations, they may also benefit from the appropriate competencies. Competencies ensure or promote quality in evaluation practice, as individuals who possess the right competences are more likely to produce high quality and utilisation-focused evaluations (Mcguire and Zorzi, 2005^[146]).

Simply put, evaluators' competencies imply having the appropriate skills, knowledge, experience and abilities (Stevahn et al., 2005^[147]) (American Evaluation Association, 2018^[148]). Nevertheless, the wide variety of contexts (internal or external evaluations) and fields (health, education, etc.) in which policy evaluations take place means that it has been difficult for literature to draw out a universal set of competencies needed for evaluators (King et al., 2001^[149]) (Stevahn et al., 2005^[147]). The knowledge, skills and abilities required to conduct policy evaluation are indeed situation dependent: depending on the policy being evaluated, the resources available, the needs of the client and stakeholders, etc. (Mcguire and Zorzi, 2005^[146]). Evaluation networks and associations have worked to establish a list of core competencies required to be an evaluator, in an effort to professionalise evaluations (Podems, 2013^[150]). The American Evaluation Association, for instance, has developed a list of core evaluator competencies (American Evaluation Association, 2015^[151]), which focus on the professional, the technical, the interpersonal, the management and organisational skills necessary to be an evaluator – thus reflecting the wide variety of competencies such a profession requires beyond technical expertise.

Regardless of their heterogeneity, OECD countries have recognised the crucial role of competencies in promoting quality evaluations. In fact, survey data shows that a majority of main respondents (17 main respondents, of which 13 OECD countries) use mechanisms to support the competence development of evaluators. Sector level practices do not differ significantly, as 16 health and 13 PSR respondents report having competence requirements for evaluators. In reality, competency requirements are the most commonly used measure to promote quality amongst respondents over all sectors (Figure 3.4).

Competency development covers a range of training and support function, aimed either at individual evaluators or at organisations in their entirety – as will be further explored in the following section.

Figure 3.4. Competence requirements for evaluators



Note: For the main institution n=42 (35 OECD member countries). For the Health ministries n=31 (28 OECD member countries). 9 countries (7 OECD member countries) did not participate on this survey. Moreover, 2 countries (1 OECD member country) are not included as they answered that none of the policies that fall in their institution's responsibility are evaluated. For the PSR ministries n=25 (20 OECD member countries). 11 countries (10 OECD member countries) did not participate on this survey. Moreover, 6 countries (5 OECD member countries) are not included as they answered that none of the policies that fall in their institution's responsibility are evaluated. Answers reflect responses to the question "How does your government ensure the quality of evaluations across government". Countries that reported no mechanisms to ensure the quality of evaluations across government are equal to zero.

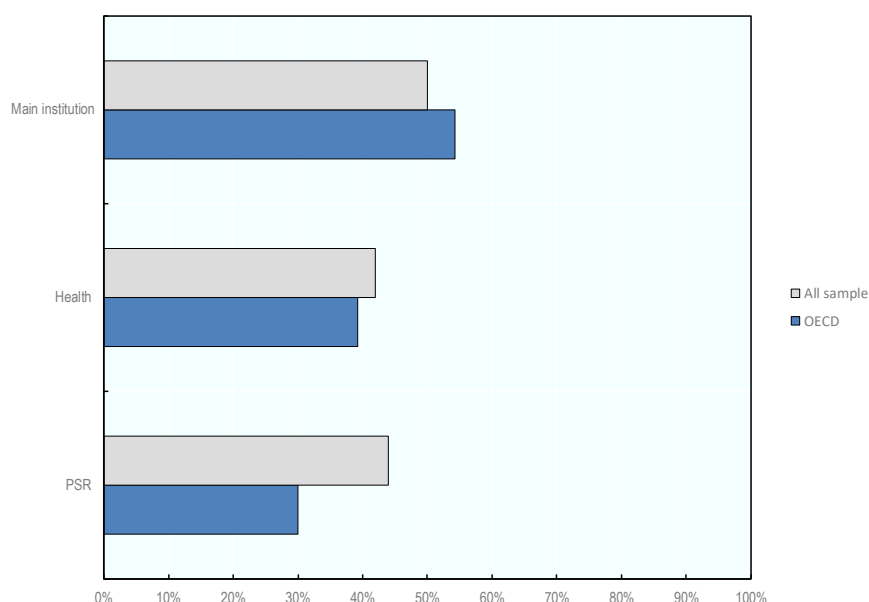
Source: OECD Survey on Policy Evaluation (2018).

Promoting individual evaluators' competencies

Training for internal or external evaluators

The appropriate competencies to carry out quality evaluations can also be developed by training internal and/or external evaluators, a mechanism that a number of surveyed countries have used. OECD survey data shows that training evaluators is the most commonly used technique for competency development: half (21) of respondent countries (including 19 OECD countries) implement such trainings. This practice is also relatively frequent at the sector level, with about 13 ministries of health and 11 ministries of Public Sector Reform organising training for their evaluators.

Figure 3.5. Training for internal or external evaluators



Note: For the main institution $n=42$ (35 OECD member countries). For the Health ministries $n=31$ (28 OECD member countries). 9 countries (7 OECD member countries) did not participate on this survey. Moreover, 2 countries (1 OECD member country) are not included as they answered that none of the policies that fall in their institution's responsibility are evaluated. For the PSR ministries $n=25$ (20 OECD member countries). 11 countries (10 OECD member countries) did not participate on this survey. Moreover, 6 countries (5 OECD member countries) are not included as they answered that none of the policies that fall in their institution's responsibility are evaluated. Answers reflect responses to the question, "How does your institution support the competence development of evaluators? (Check all that apply)". The option "Other" is not included.

Source: OECD Survey on Policy Evaluation (2018).

Evaluator training curricula may be created at the level of individual ministries (see Box 3.15 for an example for Slovakia), or homogenised across government, such as in Austria. Indeed, in Austria, several ministries such as the Ministry of Finance and the Ministry of Women and Public Services collaborated to provide a manual that gives guidance on training public officials on evaluation matters (Bundesministerium für Finanzen and Bundesministerin für Frauen und öffentlichen Dienst, 2013^[152]).

Box 3.15. Training Evaluators in Slovakia

Before entering the analytical team of a given ministry, Slovakian analysts working on policy evaluation have to pass a test that assesses their competencies in light of quality standards. For instance, some institutes use a centralised test that examines the analytical skills of the candidates for an evaluation job. Other institutes use their own tests to account for the specific evaluation requirements they have. Such requirements are good knowledge of econometric and qualitative methods and expertise on the specific policy topics the ministry focuses on.

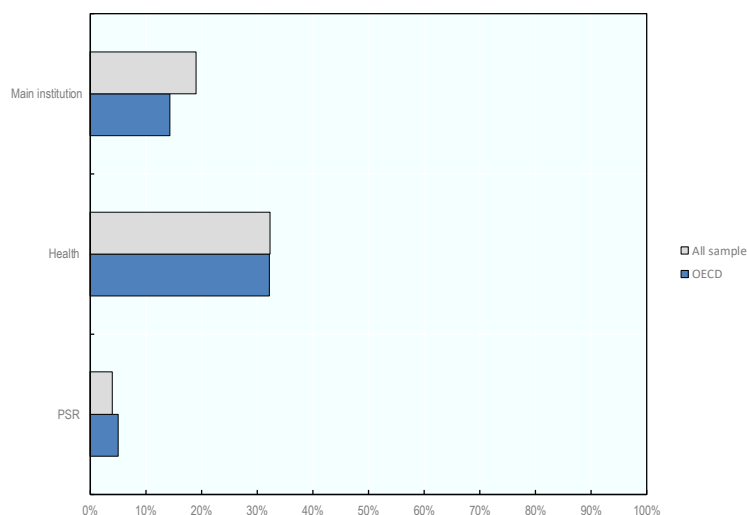
As analysts join policy evaluation units in ministries, they are offered the opportunity to attend a **broad variety of courses** to deepen their knowledge of the evaluation of a specific policy topic. For this purpose, the Value for Money Institute provides an extensive list of recommended courses. To participate, analysts have to provide documents, such as a motivation statement explaining their interest in the course. A board then reviews these documents and decides whether or not to offer the analyst a position in the course. Under the “Harvard 2 programme”, the European structural and investment funds covers the expenses for these courses.

Source: OECD (2018) Survey on Policy Evaluation.

A specific job category for evaluators in the government

A further competency development strategy that has been implemented by some governments has been to establish a specific job category for evaluators. This mechanism has been adopted by 8 main respondents, of which only 5 OECD countries. At the sector level, 10 ministries of health reported having specific evaluator positions. On the other hand, Austria is the only surveyed country that has a job category for evaluators in its Public Sector Reform Ministry.

Figure 3.6. A specific job category in government



Note: For the main institution n=42 (35 OECD member countries). For the Health ministries n=31 (28 OECD member countries). 9 countries (7 OECD member countries) did not participate on this survey. Moreover, 2 countries (1 OECD member country) are not included as they answered that none of the policies that fall in their institution's responsibility are evaluated. For the PSR ministries n=25 (20 OECD member countries). 11 countries (10 OECD member countries) did not participate on this survey. Moreover, 6 countries (5 OECD member countries) are not included as they answered that none of the policies that fall in their institution's responsibility are evaluated. Answers reflect responses to the question, “How does your institution support the competence development of evaluators? (Check all that apply)”. The option “Other” is not included..

Source: OECD Survey on Policy Evaluation (2018).

In particular, some OECD countries have developed dedicated professions, aimed at promoting policy evaluation across government. In the UK, a total of 15 000 analysts are based across the government departments. In Ireland, the Irish Government Economic and Evaluation Service (IGEES) operates as an integrated, cross-Government service, supporting better policy formulation through economic analysis and policy evaluation (IGEES, 2014^[153]). In the US, the recent Foundations for Evidence-Based Policy Making Act requires agencies to create three new positions: evaluation officer, statistical official, and chief data officer. It also requires the creation of a new (or enhancement of an existing) job series in the civil service for program evaluation.

Box 3.16. Policy evaluation as a profession in Ireland and the UK

The **Irish Government Economic and Evaluation Service** (IGEES) was created in 2012 under an initiative to extend analytic capacities for evidence-informed policy making across whole of Government. Today, the IGEES still plays a major role in building capacities to achieve better policy formulation and implementation in all Government departments on the basis of economics, statistics and evaluation practices. This service particularly aims at improving the design and targeting of policies and contribute to better outcomes for citizens by building on existing analytical work and playing a lead role in policy analysis.

IGEES staff are integrated in each department, adding their specific analytic and policy skills and expertise across whole of Government. More than 160 IGEES staff work across all of the Irish Government's Departments at different hierarchical levels, including assistant principal and administrative officer. They are either serving civil servants or staff directly recruited through the open competition process of the IGEES stream. The latter are graduates, experienced economists, evaluators and policy analysts who join analytical resources in all departments. As IGEES is an established brand in Ireland among economics graduates, this has ensured a continuous inflow of quality trained professional staff in economics across government. IGEES supports capacity building and skills enhancement and transfer for individuals and Departments through structured mobility, a learning and development framework and targeted opportunities, and platforms for discussion on analytical output and its relevance for policy. The IGEES Learning and Development (L&D) Framework intends to support capacity development according to specific individual and business needs of each Department. Following a consultation process, a cluster of skills and competencies specific to IGEES roles was developed. These skills include policy and data analysis, evaluation, quantitative methods, application of economics and civil service competencies (OECD 2020). (OECD, 2020^[98])

The UK Government Social Research Profession

The Government social research (GSR) profession is one of the civil service professions that works alongside other analysts (economists, statisticians and operational researchers). GSR professionals use the core methods of social scientific enquiry, such as surveys, controlled trials, qualitative research, case studies and analysis of administrative and statistical data in order to explain and predict social and economic phenomena for policymaking.

Members of the GSR profession come from a wide variety of social science backgrounds, including candidates with degrees in psychology, geography, sociology and criminology. The GSR profession has its own competency framework that begins with entry-level graduates as part of the fast stream to members of the senior civil service and most UK government departments would have a chief social researcher who leads and supports the activity of social researchers within the department.

Sources: OECD (2020) (OECD, 2020^[98]) *Study of the Irish Government Economic and Evaluation Service*; IGEES (2017) *Work Programme for 2018 and IGEES Achievements in 2017*; UK Government, "Government Social Research Profession". Accessed September 2nd 2019. <https://www.gov.uk/government/organisations/civil-service-government-social-research-profession/about>

Certification system for evaluators

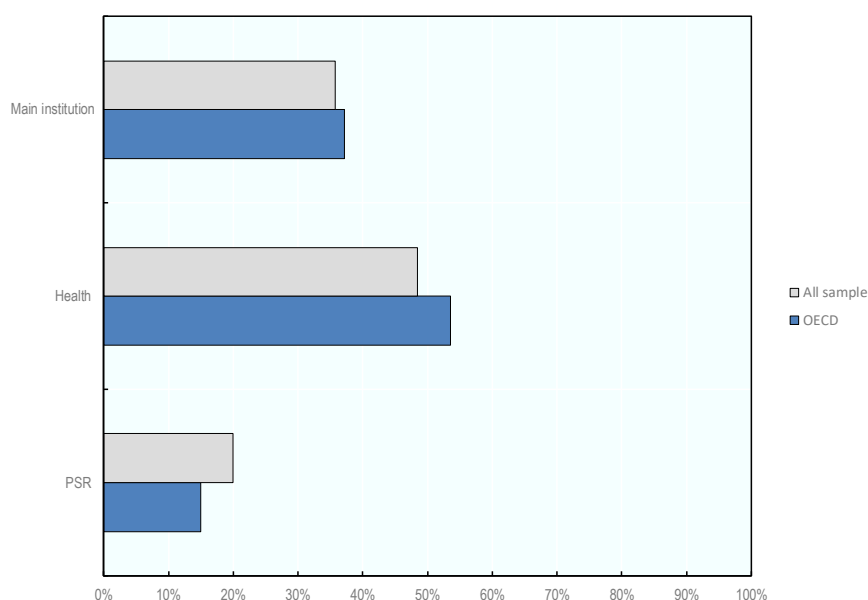
Finally, certification systems for evaluators is the least common mechanism for competency development among the countries surveyed, as out of the 42 countries surveyed, only Korea and Colombia have indicated using it.

Organisational measures for the promotion of competencies

Advisory panels and committees

Fifteen respondents also use organisational measures such as advisory panels and committees in order to promote the quality of evaluations. OECD data shows that panels and committees may be composed of either policy practitioners, managers or evaluations experts. They may be established on an *ad hoc* basis or systematically. Their main aim is to provide comments and feedback throughout the different phases of implementation of the evaluation (design, data collection, synthesis, etc.).

Figure 3.7. Advisory panels/ steering committees



Note: For the main institution n=42 (35 OECD member countries). For the Health ministries n=31 (28 OECD member countries). 9 countries (7 OECD member countries) did not participate on this survey. Moreover, 2 countries (1 OECD member country) are not included as they answered that none of the policies that fall in their institution's responsibility are evaluated. For the PSR ministries n=25 (20 OECD member countries). 11 countries (10 OECD member countries) did not participate on this survey. Moreover, 6 countries (5 OECD member countries) are not included as they answered that none of the policies that fall in their institution's responsibility are evaluated. Answers reflect responses to the question, "How does your institution support the competence development of evaluators? (Check all that apply)". The option "Other" is not included.

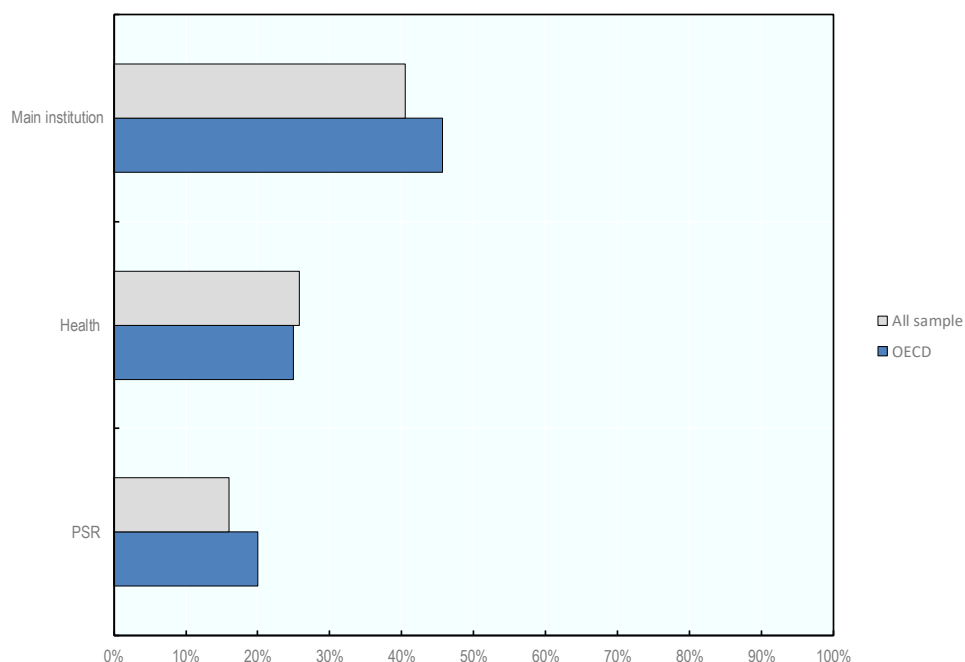
Source: OECD Survey on Policy Evaluation (2018).

Establishment and/or support a network of evaluators

OECD data shows that 17 main respondents have established or support a network of evaluators. Such networks are less common at the sector level. Examples can be found in the United States, Japan or Norway. In Norway, the EVA-forum is an informal network organisation chaired by the Agency for Financial Management. It is aimed at sharing experiences on issues regarding the evaluation initiation phase, writing terms of reference, follow-up during and after evaluation, and the sharing of evaluation

results. The forum organises several networking/workshop seminars per year and one national evaluation conference yearly which brings together over one hundred participants. The network collaborates closely with the national evaluation association, in which both Government, researchers, academics and consultants are members.

Figure 3.8. Establishment of a network of evaluators



Note: For the main institution $n=42$ (35 OECD member countries). For the Health ministries $n=31$ (28 OECD member countries). 9 countries (7 OECD member countries) did not participate on this survey. Moreover, 2 countries (1 OECD member country) are not included as they answered that none of the policies that fall in their institution's responsibility are evaluated. For the PSR ministries $n=25$ (20 OECD member countries). 11 countries (10 OECD member countries) did not participate on this survey. Moreover, 6 countries (5 OECD member countries) are not included as they answered that none of the policies that fall in their institution's responsibility are evaluated. Answers reflect responses to the question, "How does your institution support the competence development of evaluators? (Check all that apply)". The option "Other" is not included.

Source: OECD Survey on Policy Evaluation (2018)

The role of institutions and actors beyond the executive

Outside of the executive, Supreme Audit Institutions are the main institutions that play a role in promoting the quality of evaluations.

Role of SAIs in quality of evaluations and audit of the evaluation function

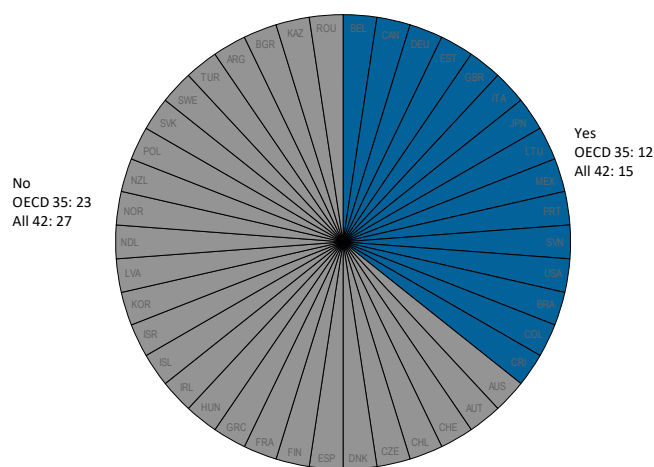
National audit institutions play an important role in evaluation discussions in the countries which have developed a more mature evaluation culture (Jacob, Speer and Furubo, 2015^[63]). Therefore, Supreme Audit Institutions (SAIs) have taken an active part in the promotion of evaluation quality. The role played by SAIs in this regard is varied, but focuses mostly on 'soft' instruments for quality assurance rather than quality control such as audits.

Firstly, SAIs are often key players in the national discourse concerning evaluation quality (see for example the role played by the United States' Government Accountability Office), bringing in their particular expertise in performance auditing, which gives countries external insights on how to improve

the quality of their evaluation systems (Jacob, Speer and Furubo, 2015^[63]). Moreover, some SAIs, such as the United Kingdom's National Audit Office, have developed guidelines for the quality of the evaluative process (National Audit Office (GBR), 2013^[154]). At the international level, the International Organisation of Supreme Audit Institutions (INTOSAI) is also supporting numerous SAIs in producing quality evaluations through the provision of specific and exhaustive guidelines (INTOSAI, 2016^[155]).

Finally, some SAIs have promoted quality evaluations by conducting audits of the national policy evaluation system. This practice is still relatively infrequent, however; OECD data shows that about 36% of countries overall (15 surveyed countries and 12 OECD countries) have seen their policy evaluation system audited by the Supreme Audit Institution in the past ten years.

Figure 3.9. Audit of the policy evaluation system by Supreme Audit Institutions



Note: n=42. Answers reflect responses to the question "Has your Supreme Audit Institution audited the executive's policy evaluation system in the past ten years".

Source: OECD Survey on Policy Evaluation (2018)

SAIs play a dual role, as they can both offer strict audits, and evaluative audits, which come closer to evaluation practices. These audits are not exclusively focused on compliance with quality norms or standards, but also look at how the system is functioning as a whole. For instance, they may assess the legal and institutional system, the evaluative processes, the information systems in place to operate them, as well as the evaluation results and their use (Operational and Evaluation Audit Division of Costa Rica, 2014^[156]). The fact that quality standards are neither explicitly included in a policy/legal framework nor in evaluation guidelines has not prevented the Belgian Court of Audit from recently analysing the performance of the national evaluation system.

Box 3.17. Audits of the evaluation system by SAIs: Examples from Belgium and Estonia

Belgium: audit of the ability of federal government departments to assess public policies

In its report to the federal Parliament, the Belgian Court of Audit examines whether federal government departments have the ability to assess public policies in an organised and professional way. An analysis of the **evaluation system's steering function** show that the majority of public services have developed evaluation practices. However, the Court underlines the lack of a central vision and strategy on evaluation within the public function, which blurs the division of labor and hindered coordination between actors. The report recommends fully incorporating evaluation in the policy cycle and budget by delivering yearly evaluation notes to Parliament. The audit also analyses the **resources dedicated to evaluation**, warning against the diminishing budget of certain services. It notes the lack of a clear evaluation function and advocates for the need to facilitate data access and develop analytic tools.

The **implementation of evaluations and quality assurance** is audited as well. According to the Court, the country's public services rarely have well-defined tasks, processes and methodologies for quality assurance. The audit, thus, reiterates the importance of quality methods and of making policies evaluable by clearly and explicitly defining the logic of intervention and collecting the necessary data. Finally, the Court characterizes the **use of evaluation results** as insufficient because public services rarely make them publicly accessible. There is also a lack of transparency, which suggests that evaluations are not seen as a means to justify public policies on the federal level.

Estonia: audit of the planning, conduct and use of impact evaluations

The Estonian National Audit Office's 2011 report to Parliament assesses the **planning and conduct of evaluations** and whether their **results** are continuously provided to Parliament and the public through coordination mechanisms. According to the audit, ministries lack coordination mechanisms and requirements for establishing evaluations. The lack of resources and capacity building dedicated to evaluative practices and the perceptions regarding evaluations are identified as reasons for the low quantity and quality of evaluations. The report recommends establishing sustainable quality control at the Executive and Parliament levels, clarifying the scope and methods of evaluations, involving stakeholders in legislative drafting, and clearly communicating impact analysis in explanatory memorandum (Estonian National Audit Office, 2011^[157]).

Sources: Belgium Supreme Court of Audit (2018); Estonia National Audit Office (2011), *The state of affairs with legislative impact assessment*

Only Slovenia, Brazil and Colombia specifically mentioned that their SAI had conducted several audits of the policy evaluation system in the past ten years. No country surveyed by the OECD has reported conducting systematic audits, as is the case with the European Court of Auditors (see Box 3.18).

Box 3.18. The role of the European Court of Auditors in auditing the evaluation system of a Directorate General

The European Court of Auditors' major activity consists in conducting performance audits, which entails examining the quality of the evaluation system of a directorate general (DG) according to European Commission standards. Such a quality evaluation system is one that ensures effective evaluation demand management, quality of supply and use of results. DGs are mandated to implement those standards to foster quality evaluation systems.

Accordingly, the European Court of Auditors offers practical guidelines to support the assessment of these evaluation systems' quality. Since evaluation systems have to be tailored to their environment, the guidelines advise thorough understanding of policies' intervention logic, their legal framework, available resources, etc. This also requires processes for programming, monitoring, supporting and reporting on evaluations.

- In terms of effectively managing evaluation demand, DGs should attach sufficient importance to the evaluation itself, which requires gaining support from high level decision-makers and creating the right incentives for carrying out evaluations.
- On the other hand, supporting evaluation quality requires DGs to implement procedures for training evaluation staff appropriately, involving stakeholders, ensuring robust methods and rigorously planning evaluations.

Finally, to foster use of evaluation results, arrangements should be set to identify users, understand their needs, communicate the results clearly and deliver them on time, and follow-up on their ultimate use.

Source: (European Court of Auditors, 2013^[158])

National evaluation associations or societies

Outside of government, national associations of evaluators play an important role in promoting the competencies of evaluators and the quality of evaluations (Cooksy and Mark, 2012^[159]). All OECD countries have a national evaluation association. Evaluation associations use a variety of approaches for encouraging competencies for quality analysis in the evaluation community. Some evaluation societies, such as the American Evaluation Association, seek to create a policy environment for quality evaluations by advocating the utility of good analysis to policy makers, establishing guidelines and increasing awareness through workshops, trainings, webinars (Cooksy and Mark, 2012^[159]). Others, such as the Canadian Evaluation Society, have developed a professional designations program, which imposes a minimum competency requirement to be considered an evaluator.

Table 3.5. National Evaluation Societies in OECD countries

Country	Name of the Society Network	Website
Australia	Australian Evaluation Society	https://www.aes.asn.au/
Austria	Austrian-German Evaluation Association	https://www.degeval.org/home/
Belgium	Flemish Evaluation Platform	http://www.evaluatieplatform.be/VEP/index.htm
Canada	Canadian Evaluation Society	https://evaluationcanada.ca/
Chile	Red Chilena de Evaluación	http://www.evaluacionpoliticaspublicas.com/
Czech Republic	Czech Evaluation Society	https://czecheval.cz/
Denmark	Danish Evaluation Society	http://danskevalueringsselskab.dk/
Estonia	Estonian Evaluation Society	http://www.praxis.ee/vana/index.php?id=1029.html
Finland	Finnish Evaluation Society	http://www.sayfes.fi/in-english/
France	French Évaluation Society (SFE)	http://www.sfe-asso.fr/
Germany	German Evaluation Society (DeGEval)	https://www.degeval.org/en/home/
Greece	Hellenic Evaluation Society	http://www.hellenicevaluation.org/index.php/el/
Hungary	Hungarian Evaluation Society	https://www.europeanevaluation.org/content/hungarian-evaluation-society
Iceland		
Ireland	Irish Evaluation Network	https://www.dcu.ie/eqi/en/index.shtml
Israel	Israeli Association for Program Evaluation	http://www.iape.org.il/en_index.asp
Italy	Italian Evaluation Association	http://valutazioneitaliana.eu/
Japan	Japan Evaluation Society (JES)	http://evaluationjp.org/english/index.html
Korea	Korean Evaluation Association	http://www.valuation.or.kr/
Latvia	Latvian Evaluation Society (LATES)	http://www.izvertesana.lv/en/about-us/
Lithuania		
Luxembourg	Luxembourg Evaluation and Foresight Society	http://solep.lu/
Mexico	Academia Nacional de Evaluadores Mexicanos (ACEVAL)	http://aceval.org/
Netherlands	Dutch Evaluation Society (VIDE)	https://www.videnet.nl/
New Zealand	Aotearoa New Zealand Evaluation Association (ANZEA)	https://www.anzea.org.nz/
Norway	Norwegian Evaluation Society	http://norskevalueringssforening.no/
Poland	Polish Evaluation Society	http://pte.org.pl/
Portugal	Portugal Evaluation Association (AvalPortugal)	https://avalportugal.wordpress.com/
Slovak Republic	Slovak Evaluation Society (SES)	http://www.evaluacia.sk/en/
Slovenia	Slovenian Evaluation Society	https://www.sdeval.si/eng
Spain	Spanish Public Policy Evaluation Society (SEE)	http://www.sociedadevaluacion.org/
Sweden	Swedish Evaluation Society	http://svuf.nu/
Switzerland	Swiss Evaluation Society (SEVAL)	https://www.seval.ch/
Turkey	Turkish Monitoring and Evaluation Society (TMES)	https://www.ived.org.tr/english
United Kingdom	UK Evaluation Society	https://www.evaluation.org.uk/about-us/
United States	American Evaluation Association	https://www.eval.org/

Ensuring evaluation for impact: promoting the use by politicians, practitioners and citizens

Understanding the use of evaluations

As policy-makers invest public funds in evaluations, their use and the ability to improve policy, programs, or projects, are key to their success. One of the most fundamental rationale for conducting policy evaluations is their usefulness in informing policy and decision-making, in general, and improving the intervention they consider, specifically. Indeed, one of the principal goals of evaluation is to support decision-making with useful insights on public issues and evidence on the impact of policies and their underlying change mechanisms.

The literature on policy evaluation use defines the concept of use threefold (Weiss and Weiaa Harvard, 1998^[160]) (Alkin and Taut, 2002^[161]) (Fleischer and Christie, 2009^[162]) (Ledermann, 2012^[163]):

- *Symbolic use* (also known as persuasive) occurs when the results of evaluations are taken up to justify or legitimise a pre-existing position, without changing it. Examples of this are when ministers use evaluations to justify their policy choices or when congressional members use findings from an evaluation in order to push for a proposition of law (Ledermann, 2012^[163]).
- *Conceptual use* happens when evaluation results lead to an improved understanding or a change in the conception of the subject of evaluation. An example of this is the identification of collateral impact of a policy or reverse causation (Ledermann, 2012^[163]).
- *Instrumental use* is when evaluation recommendations inform decision making and lead actual change in the policy being evaluated. An example of this is the reallocation of funds after a poor performance (Ledermann, 2012^[163]).

The users of evaluations include not only decision makers, for whom conceptual and instrumental use are key, but also civil servants, experts and practitioners (local authorities, programme managers, health practitioners, etc.), who are looking for increased accountability, learning and better strategic decision-making. Evaluations can be used to improve regulations, inform resource allocations on the ground or monitor the implementation of policies, etc.

Regardless of these many potential users, the use of evaluations remains a constant challenge and often falls under expectations. Despite the potential for policies to be based on evidence, in reality an effective connection with many types of research evidence in policy making remains elusive (Newman, Cherney and Head, 2017^[164]). For example, USA estimates show that under the two Obama administrations, only 1% of government funding was informed by evidence (Bridgeland and Orszag, 2013^[165]).

Furthermore, while many factors contribute to evaluation use, the role of barriers and facilitators to evidence use will vary depending on the context. The way in which specific barriers and facilitators operate, and how they interact with each other, depends on the local context. Use of evaluation is “more of an art than a science” (Results for America, 2017^[166]). Thus, in order to promote the use of evaluations, it is important to understand these determinants and their interactions, before discussing the range of practices promoted by countries to promote use.

Overview of mechanisms to promote the use of evaluations

A large majority of countries (31 countries, of which 27 OECD countries) have put in place one or several mechanisms in order to influence these determinants – and thus promote the use of evaluations (Table 3.6 next page). In general, countries have sought to promote the use of evaluations by:

- conducting utilisation-focused evaluative processes
- promoting access to evaluations
- supporting the uptake of evaluations results
- increasing demand for evaluations through competency development
- embedding use in the institutional set-up, within and outside of the executive.

Conducting utilisation-focused evaluations

Countries have developed mechanisms to ensure that evaluative processes are utilisation-focused, meaning that evaluations are conducted in a way that is fit for purpose and takes into account the needs of their primary users and the types of intended uses (Patton, 1978^[102]). Empirical research (Johnson et al., 2009^[167]) has found that user-focused evaluations share several features:

- They are methodologically robust and credible (for a discussion of determinants of credible evaluations, see the section on ‘Mechanisms to promote quality evaluations’ as well as the OECD (forthcoming^[106]). report on Principals and Standards for Good Governance of Evidence).
- Users and stakeholders are involved in the evaluation process.
- The evaluation methodology is perceived as appropriate by users.

Involving stakeholders throughout the evaluative process

Governments overall are increasingly eager to engage a wide range of internal and external stakeholders in the decision-making process to generate a broader consensus and increase the legitimacy of public-policy decisions (OECD, 2016^[43]). There is a general consensus in the academic literature that engagement with those concerned and affected by evaluations is fundamental to improving the design, relevance, transparency and, *in fine*, use (Patton, 1978^[102]) (Kusters et al., 2011^[125]) (Gauthier, 2015^[168]). Concordantly, OECD data shows that 72% of countries overall (and 71% of OECD countries) report engaging stakeholders in the evaluation of their policy priorities.

Evidence shows that policy-makers are more likely to seek and use evaluation results obtained from trusted familiar individuals or organisations rather than from formal sources (Oliver et al., 2015^[169]; Haynes et al., 2012^[170]). Stakeholder participation and interaction in the evaluative process can help build trusted relationships and increase the opportunities for evaluation results to impact policy making. Similarly, communicating findings to stakeholders as the evaluation progresses, or involving stakeholders in the design of the evaluation, can favour their adherence and understanding of the results (Fleischer and Christie, 2009^[162]).

Table 3.6. Mechanisms to promote the use of evaluations

	Management response mechanism in place	Incorporation of findings into the budget cycle	A rating / grading system	Coordination platform to promote use of evidence	Discussions of findings at the Council of Ministers	No specific initiatives in place
Australia	○	●	○	○	○	○
Austria	●	●	○	○	○	○
Belgium	○	○	○	○	○	●
Canada	●	●	○	●	○	○
Chile	○	●	○	○	●	○
Czech Republic	○	○	○	○	○	●
Denmark	○	○	○	○	○	○
Estonia	●	●	○	●	○	○
Finland	○	●	○	●	●	○
France	○	●	○	○	○	●
Germany	○	●	○	●	●	○
Great Britain	○	●	○	●	○	○
Greece	●	●	○	○	●	○
Hungary	○	○	○	○	●	○
Iceland	○	○	○	○	○	●
Ireland	○	●	○	●	○	○
Israel	○	●	○	○	○	○
Italy	○	○	○	○	○	○
Japan	●	●	●	●	●	○
Korea	●	○	○	○	●	○
Latvia	●	●	○	●	●	○
Lithuania	○	●	○	○	●	○
Mexico	●	●	●	●	●	○
Netherlands	○	●	○	○	○	○
New Zealand	●	○	○	○	○	○
Norway	○	○	○	●	○	○
Poland	○	○	●	●	○	○
Portugal	○	●	○	○	●	○
Slovakia	●	●	○	○	○	○
Slovenia	○	○	○	○	○	●
Spain	○	○	○	○	○	●
Sweden	○	●	○	○	○	○
Switzerland	○	○	○	○	○	●
Turkey	○	○	○	○	○	●
United States	○	●	○	●	○	○
OECD Total						
● Yes	10	21	3	12	11	8
○ No	25	14	32	23	24	27
Argentina	○	○	○	○	●	○
Bulgaria	○	○	○	○	○	●
Brazil	○	○	○	●	○	○
Colombia	○	○	○	●	●	○
Costa Rica	●	○	●	○	●	○
Kazakhstan	○	○	○	○	○	●
Romania	○	○	○	○	○	●

Note: n=42 (35 OECD member countries). Answers reflect responses to the question "How does your government promote the use of the findings of policy evaluations". The option "Other" is not included. A rating/grading system refers to classify the robustness of evidence provided and recommendations derived from the policy evaluations exists.

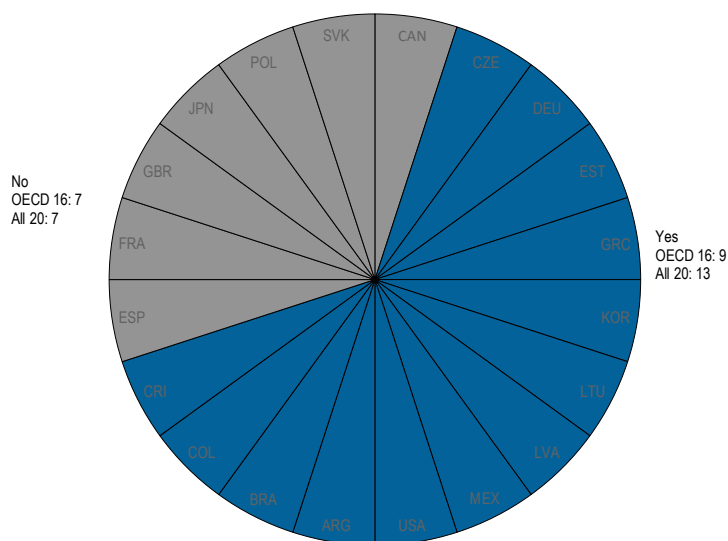
Source: OECD Survey on Policy Evaluation (2018)

Requirements for stakeholder participation

In order to promote stakeholder participation, 65% of all countries (56% of OECD countries) have adopted formal requirements for stakeholder engagement in their legal/policy framework related to policy evaluations.

In the Netherlands, for example, the Ministry of Finance's *Regulations for periodic evaluation research* (15 March 2018) lays down rules for the participation of stakeholders in periodic evaluations. With each policy evaluation, at least one independent expert must give an opinion on the quality of the evaluation. Similarly, the European Commission's *Better Regulation Guidelines* contain a chapter that describes standards for stakeholder engagement. According to these guidelines, views from stakeholders should be included in the evaluation of all programmes and policies issued by the Commission, as well as initiatives with impact assessments (European Commission, 2017^[171]). Public participation in the processes for designing new regulations are also underlined in the area of regulatory policy [GOV/RPC(2019)].

Figure 3.10. Requirements related to stakeholder engagement in policy/legal frameworks



Note: n=20 (all countries surveyed who have a policy framework, among which 14 are OECD). Answers reflect responses to the question, "Which element(s) do(es) the document/s referred to under Q4 and Q5 cover concerning policy evaluation across government? (Check all that apply)".

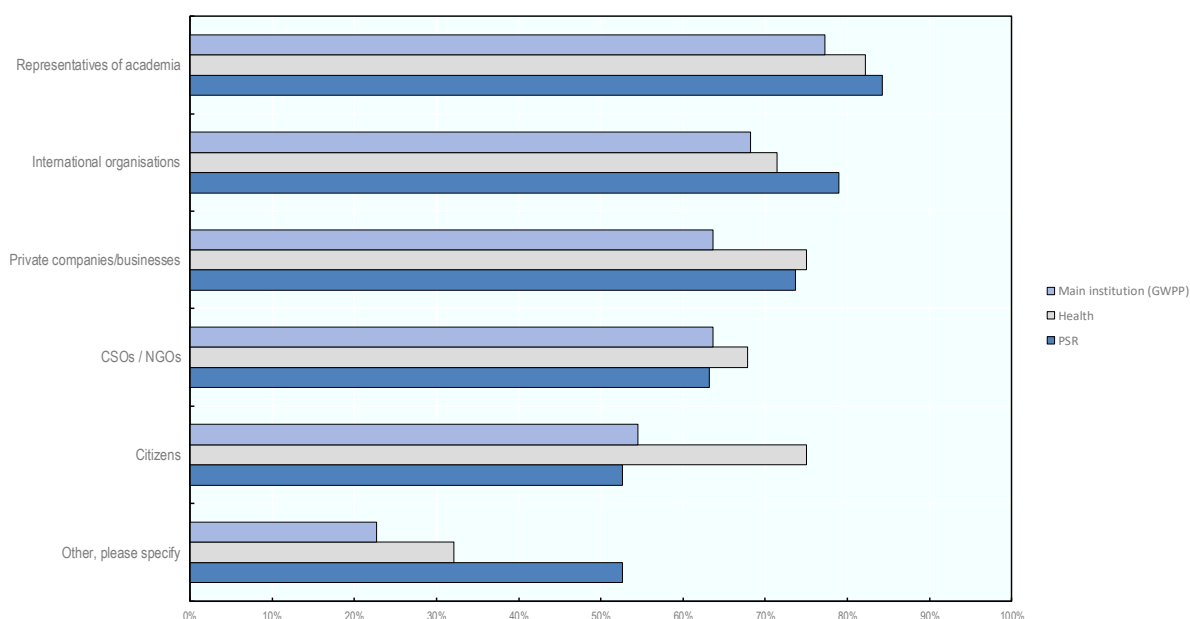
Source: OECD Survey on Policy Evaluation (2018).

Variety of stakeholders involved in the evaluation process

According to the OECD survey, stakeholders include a variety of actors such as citizens, CSOs/NGOs, representatives of academia, representatives of the private sector and international organisations. Still, it is very revealing that overall representatives of the academia (17 countries for the evaluation of government-wide policy priorities (GWPP), 23 for that of health sector policies, and 16 for PSR) and the private sector (14 GWPP, 21 health, 14 PSR) are more likely to be engaged in the evaluation process than citizens according to the survey results.

These results suggest that countries mostly engage with stakeholders emanating from traditional sources of authority and expertise (academia, international organisations, the private sector). Yet, citizens, as the primary intended users of the policy being evaluated, can be considered to be the most important stakeholders to include in the evaluative process (Kusters et al., 2011^[125]). Other stakeholders that are not represented include staff and managers, who can produce actionable knowledge by being engaged in the evaluation process (Gauthier, 2015^[168]). In an era of relative discontent with public policies in a significant number of countries, this apparent gap of public engagement may reveal the need to explore how to reengage with citizens on evaluation results that they will both understand and find useful. This is particularly true for key challenges such as taxation, health or climate change that generate significant concerns among citizens.

Figure 3.11. Types of stakeholders engaged in policy evaluations



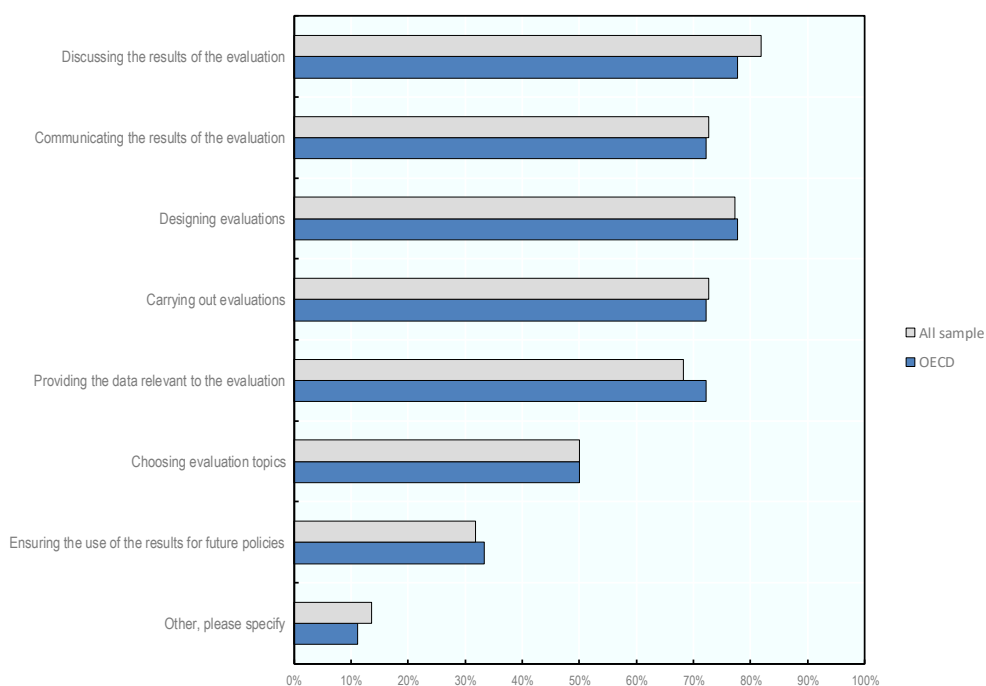
Note: For the main institution in charge of government-wide policy evaluation $n=21$. 4 countries answered that they do not have government-wide policy priorities. Moreover, 9 countries answered that they do not evaluate their government-wide policy priorities. 8 countries answered that they do not engage stakeholders in the policy evaluation process. Health ministries $n=28$. 9 countries did not participate on this survey. Moreover, 2 countries are not included as they answered that none of the policies that fall in their institution's responsibility are evaluated. 3 countries answered that they do not engage stakeholders in the policy evaluation process. For the PSR ministries $n=19$. 11 countries did not participate on this survey. Moreover, 6 countries are not included as they answered that none of the policies that fall in their institution's responsibility are evaluated. 5 countries answered that they do not engage stakeholders in the policy evaluation process. Answers reflect responses to the question "Which stakeholders are engaged in the evaluation of government-wide policy priorities (Check all that apply)".
Source: OECD Survey on Policy Evaluation (2018)

Engagement at different stages of the evaluation process

Involving stakeholders during every step of the evaluation can be very useful. The earlier and more actively the intended users are involved in an evaluation process and with dissemination of results, the more likely they are to use the evaluation's results (Patton, 1978^[102]). However, there is no general agreement in literature on the recommended degree of engagement between the evaluator and the users, as well as when those should be involved in the evaluation process (Fleischer and Christie, 2009^[162]).

Country practices are mixed. A large majority of respondents engage stakeholders during the implementation of the evaluation (designing evaluations, providing the data relevant to the evaluation, carrying out evaluations) and during the dissemination of the results (discussing the results of the evaluation, communicating the results of the evaluation). Fewer countries engage stakeholders when deciding what policies should be evaluated and in following-up on the use of results.

Figure 3.12. Stakeholder engagement in the evaluative process



Note: For the main institution in charge of government-wide policy evaluation $n=21$ (15 OECD member countries). 4 countries (all OECD member countries) answered that they do not have government-wide policy priorities. Moreover, 9 countries (7 OECD member countries) answered that they do not evaluate their government-wide policy priorities. 8 countries (7 OECD member countries) answered that they do not engage stakeholders in the policy evaluation process.

Canada offers an interesting example of stakeholder involvement at the early stages of an evaluation process. Results in some sectors are less clear-cut. Overall, results may suggest that stakeholder engagement in the field of policy evaluation may be aimed at symbolic use, whereby evaluations are conducted in order to justify prior decisions. Yet, stakeholders' willingness to consider evaluations evidence is key to promote learning. Similar results can be found in the *OECD Regulatory Policy Outlook* (OECD, 2018^[172]), which found that most consultation efforts in regards to regulatory policy development continue to focus on later stages of the rule-making process, i.e. when a preferred solution has been identified and/or a draft regulation been prepared.

Box 3.19. The experience of Alberta in Canada and the “What We Heard” report

The Canadian province of Alberta’s ministry of health has established an innovative approach to communicating the results of a citizen-centred approach of policy evaluations, especially its inputs, the methodology used, as well as the outputs. In 2005, the Government of Alberta introduced the “Getting on with Better Health Care” package, which contained 13 concrete actions for the advancement of the health care system. One of these actions included the “Health Policy Framework”. In order to ensure a needs-tailored design and implementation of this framework, the government inquired the opinion of 420 health system stakeholders, health care professionals, unions, municipal leaders, educators and community organisations. Through letters and e-mails, meetings, phone calls and online expression of opinions, the government heard from 4 056 individuals from Alberta and their suggestions on how to design and implement the best health policy framework possible.

The consultations were collected and summarised in a consultation report entitled, “What We Heard...from Albertans during March 2006”, and is accessible to everyone on Alberta’s Ministry for Health’s website. In a concise and easy to read report, the ministry of health provides information on the approach of the consultation, the timeframe, the content of the input received and the “lessons learned” from the consultation process. These results of the evaluation process were subsequently used to further improve approaches to integrate citizens’ opinions. The necessity and value added of such consultation processes is underscored by one of the major findings of the report, “Albertans would like more information and communication about Alberta’s policy directions to better understand the framework and what it will mean for them.”

Source: based on OECD (2016), *Open Government: The Global context and the Way Forward*; Alberta Health and Wellness (2006), “What We Heard... from Albertans during March 2006”, www.health.alberta.ca/documents/What-We-Heard-Report-2006.pdf (accessed 02 August 2016).

Designing and implementing evaluations for use

The evaluation’s set-up, understood as the planning, resources and communication channels involved, also affects the use of evaluations in policy-making. The set-up needs to be tailored to the policy maker’s needs if use is to be facilitated in practice. The resources for evidence should match the demand of policy makers in terms of timing and format. Finally, the evaluation questions foreseen by the evaluator should be set to match the users’ needs (Patton, 1978^[102]).

Box 3.20. The concept of Utilization-Focused Evaluations

The concept of *Utilization-focused Evaluations* (Patton, 1978^[102]), developed by Michael Quinn Patton, refers to the principle according to which an evaluation should be useful to its intended users and should be judged based on its utility. Evaluations should be planned and conducted in ways that enhance their likely use. The following 17-step guidelines were identified to increase the impact and usefulness of evaluations:

1. Assess and build program and organisational readiness for utilisation-focused evaluation
2. Assess and enhance evaluator readiness and competence to undertake a utilization-focused evaluation
3. Identify, organise and engage primary intended users: the personal factor
4. Situation analysis conducted jointly with primary intended users
5. Identify and prioritise primary intended uses by determining priority purposes
6. Consider and build in process uses if and as appropriate
7. Focus priority evaluation questions
8. Check that fundamental areas for evaluation inquiry are being adequately addressed: implementation, outcomes and attribution questions
9. Determine what intervention model or theory of change is being evaluated
10. Negotiate appropriate methods to generate credible findings that support intended use by intended users
11. Make sure intended users understand potential methods controversies and their implications
12. Simulate use of findings: evaluation's equivalent of a dress rehearsal
13. Gather data with ongoing attention to use
14. Organize and present the data for interpretation and use by primary intended users: analysis, interpretation, judgment, and recommendations
15. Prepare an evaluation report to facilitate use and disseminate significant findings to expand influence
16. Follow up with primary intended users to facilitate and enhance use
17. Meta-evaluation of use: be accountable, learn, and improve.

Sources: Patton (1978), *Utilization-focused evaluation*, OECD (2019), *Evaluating Public Sector Innovation Support or hindrance to innovation?*

Methods and tools to promote access to evaluation results

Policy makers and stakeholders cannot use evidence and the results of evaluation if they do not know about it (Haynes et al., 2018^[173]). The first step to promote use is therefore that the results be made available to their intended users – simply put, that they be communicated and disseminated to stakeholders. While communication supplies evidence to specific users and publics of policy evaluations, evidence dissemination aims to maximise general access to research and increase stakeholders' understanding of, and confidence in, such content.

In many countries, unpublished evaluations may be sought through freedom of information laws. Freedom of information laws (FOI) – also referred to as access to information laws – presume a principle of maximum disclosure of information, i.e. the information held by the state is in principle available to the public. However, these laws also contain a list of exemptions that may be applied to justify withholding certain information from disclosure (see the OECD's *Government at a Glance* (OECD, 2011^[174]) for an overview of an overview of FOI in OECD countries). It is worth noting that in European countries, evaluations of cohesion policies are made publicly available, but that may not be the case for all evaluations done at the national level, depending on the countries.

Nevertheless, the figure above shows an uneven level of dissemination and publicity of the results, with much scope to reflect on the potential to further increase use through greater awareness and use of modern communication and dissemination techniques. This was for example highlighted as a key message of a recent OECD study of the Irish Government Economic Evaluation System. (OECD, 2020^[98])

Evaluation databases and portals

Online databases, for instance, seek to increase access to specific types of research. Some surveyed countries have created national databases or evaluation portals with the aim of centralising evaluation evidence in one easily accessible place.

Box 3.21. Evaluation portals to promote the use of evidence

Poland's national evaluation database for the evaluation of cohesion policy

All evaluations commissioned in Poland, including those concerning the implementation of EU funds, must be made accessible to the public. Concerning the evaluations related to Cohesion Policy, a national database has been created: all evaluations are published on the website www.ewaluacja.gov.pl. This platform shares the results of more than a thousand studies conducted since 2004, as well as methodological tools aimed at evaluators.

Norway's evaluation portal

Norway's evaluation portal (<https://evalueringsportalen.no/>) is a publicly accessible web service that gathers all the findings of evaluations carried out by the central government. This database is operated by the Directorate for Financial Management and the National Library of Norway. It contains evaluations carried out on behalf of government agencies from 2005 until today, as well as a selection of central evaluations from 1994 to 2004. Evaluation reports are registered in the database as soon as they are made available to the public. Moreover, the portal provides evaluation guidelines, a calendar of the key activities in the evaluation area, news and professional papers.

By increasing accessibility to evaluation results, the portal allows the use and reuse of the knowledge and findings from evaluations in all state policy areas, in future evaluations and in society as a whole. It ultimately allows increased legitimacy and transparency regarding government activities.

Source: OECD Survey on Policy Evaluation (2018).

Communication strategies and tools

Research suggests that in isolation, publicity alone does not significantly improve uptake of evaluations in policy-making (Haynes et al., 2018^[173]; Langer, Tripney and Gough, 2016^[175]; Dobbins et al., 2009^[176]). Rather, the way evidence is presented should be strategic and driven by the evaluation's purpose and the information needs of intended users (Patton, 1978^[102]). When evaluation results are well synthesised, tailored for specific users and sent directly to them, their use is facilitated (Haynes et al., 2018^[173]). Tailored communication and dissemination strategies that increase access to clearly presented research findings are very important for use.

These strategies can include use of infographics, tailored synthesis of research evidence, for example in the form of executive summaries, dissemination of 'information nuggets' through social media, seminars to present research findings, etc. (OECD, 2016^[144]) (OECD, 2020^[99]). The UK What Works Centre – including Education Endowment Foundation, the Early Intervention and the What Works Centre for Local Economic Growth – produce a range of policy briefs to disseminate key messages to its target audience. Similarly, in Canada, departments are diffusing evaluation findings beyond departmental websites via such platforms as Twitter and LinkedIn.

Methods for reviewing and assessing the evidence base

Several methods exist for reviewing and assessing the evidence base. Portals serving as passive repositories of information are less likely to promote evidence use (Results for America, 2017^[166]). Compiling evaluations in portals or databases runs the risk of information overload, thus hindering the incorporation of findings and reducing the effectiveness of evaluation. As the number of evaluations increases, it becomes more difficult for policy makers and practitioners to keep abreast of the literature. Yet, policies should ideally be based taking into account the full assessment drawn from the body of evidence, not single studies, which may not provide a full picture of the effectiveness of a policy or programme. In addition, such repositories do not necessarily allow stakeholders to understand the quality of the evidence produced by an evaluation: its rigor or replicability for example.

These needs have led to an increase in the use of evidence synthesis. Evidence syntheses, through secondary processing of existing evaluations, provide a vital tool for policy makers and practitioners to:

- inform them about what can be known, or derived, from previous research
- understand what works and how it works.

Evidence synthesis methodologies seek to not only aggregate evaluation findings and review them in a more or less systematic manner (for a discussion of methods), but also assess and rate the strength of the evidence. Evidence syntheses provide a useful dissemination tool since they allow decision-makers to access large bodies of evidence, as well as rapidly assess the extent to which they can trust it. They can also play an important role in promoting the quality of evaluations, as discussed in the section on 'Mechanisms to promote the quality of evaluations'.

Box 3.22. Different methodologies for reviewing the evidence base

Effective policy-making requires using the best available evidence, which itself requires reviewing and choosing from the already existing evidence on the policy question. Different reviewing methods enable managing and interpreting the results of a large evidence base:

- **Quick Scoping Review:** this non-systematic method can take from 1 week to 2 months. It consists in doing a quick overview of the available research on a specific topic to determine the range of existing studies on the topic. It allows mapping the literature concerning a delimited question by using only easily accessible, electronic and key resources, going up to two bibliographical references.
- **Rapid Evidence Assessment (REA):** this systematic and more time-consuming method (2 to 6 months) consists in quickly overviewing the existing research on a specific policy issue and synthesising the evidence provided by this research. It intends to rigorously and explicitly search and critically appraise this evidence. To gain time, it may limit certain aspects of the systematic review process, such as narrowing the REA question or the type and breadth of data considered. Shortening the traditional systematic review process provides a rapid synthesis of the existing relevant evidence, but risks introducing bias.
- **Systematic Review:** this is the most robust method for reviewing, synthesising and mapping existing evidence on a particular policy topic. It is more resource-intensive, as it can take up to 8 to 12 months minimum and requires a research team. It has explicit objectives and a thorough search strategy that considers a broad range of data. Studies are chosen and screened according to explicit and uniform criteria, and reasons for excluding certain studies have to be stated. This transparent and comprehensive method maximally reduces bias in the search, choice and synthesis of the existing research. Moreover, it allows the creation of a cumulative and sound evidence base on a specific policy subject. Lastly, systematic reviews are applicable to quantitative studies as well as other types of questions.

Source: The UK Civil Service, *What is a Rapid Evidence Assessment?*
<https://webarchive.nationalarchives.gov.uk/20140402163359/http://www.civilservice.gov.uk/networks/gsr/resources-and-guidance/rapid-evidence-assessment/what-is> (Accessed August 12th 2019).

According to the survey results, only a small number of countries conduct evidence synthesis within government. In fact, only two countries (Japan and Poland) declared using a rating system to classify the robustness of evaluations at the level of the main institution in charge of policy evaluations. The results from the UK What Works Centres, or corresponding knowledge brokerage institutions in the US or Australia, may not have been included as they might not be considered as being “within government”.

These clearinghouses or what works centres play a significant role as knowledge brokers. According to the Results for America initiative, clearinghouses can be defined as “an information source that aggregates, standardizes, reviews and rates the evidence base of interventions” (Neuhoff et al., 2015^[177]) (Results for America, 2017^[166]). Clearinghouses, therefore, conduct evidence syntheses to make information available to decision-makers and translate the research into language relevant to them.

Because of the nature of their mandate, clearinghouses usually work at arms’ length of government (for instance receiving government funding but functioning autonomously, such as the What Works Network in the UK) or completely independently from government (in the case of nongovernment initiatives, such as the Campbell Collaboration or the Cochrane Library, for example). Many focus on only one area of specialisation (Results for America, 2017^[166]) and have their own review and rating process.

Box 3.23. The United Kingdom's What Works Network

The What Works Network comprises seven independent What Works Centres and two affiliate members. It intends to support the government and other organisations in creating, sharing and using high quality evidence to make better decisions for the improvement of public services. What Works is a unique national approach taken by a government to inform decision-making by the best available or created evidence. Its success can be associated to its three key features:

Its autonomy:

- a. The Network operates at arm's length from government and independently assesses the evidence that it encourages policymakers to incorporate in their decisions.
- b. It is also funded out of non budgetary resources, such as lottery funds and its operations concern the public sector at national and local levels, covering policy areas that receive more than £200 billion of public spending.

Its role in promoting use of evidence:

- c. It allows policy makers, commissioners and practitioners to use evidence on what works to make decisions and provide cost-efficient and useful services, distinguishing itself from standard research centres.
- d. Several centres also support the development of a civil service with the skills, capability and commitment to use evidence effectively (Results for America, 2017^[166]).
- e. Additionally, the What Works National Adviser located in the Cabinet Office runs a Cross-Government Trial Advice Panel, which includes experts from academia and government who provide free support to all civil servants to assess whether policies are working.
- f. This Adviser also frames findings from all Centres in an accessible and understandable format and shares them across government. This practice encourages cross-government discussions on 'what works' and assists policy makers in making evidence-based decisions regarding investment in value for money services that are intended to have a positive impact on citizens

Its role in producing systematic reviews:

- g. Where evidence is lacking, centres create high quality synthesis reports in their policy domain (What Work Centres are usually focused on one policy area, such as wellbeing or early intervention).
- h. Centres also collate existing evidence on the effectiveness of policies and practices.

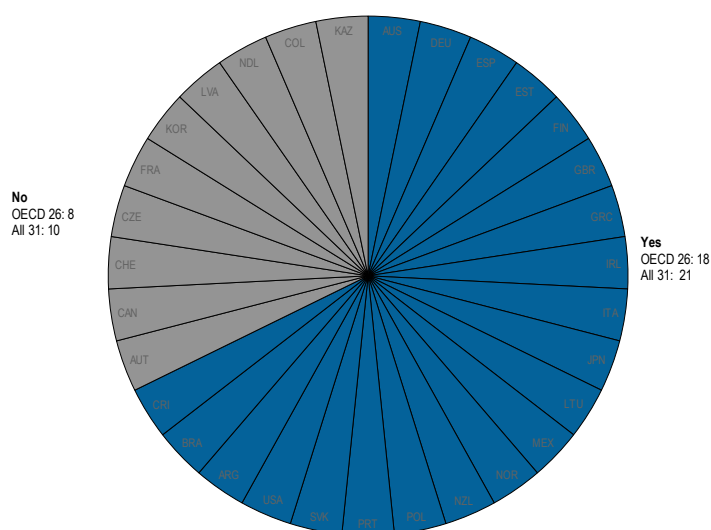
Sources: UK Government "What Works Network" <https://www.gov.uk/guidance/what-works-network> (Accessed September 2nd 2019), (Results for America, 2017^[166]).

Guidelines to promote the uptake of evaluation results

While communication strategies, platforms and evidence synthesis methodologies are designed to give users quick and easy access to clear evaluation evidence, they do not systematically translate to better uptake of policy evaluations in decision-making.

According to the OECD data, 21 countries (of which 18 OECD members) have developed guidelines on public policy evaluations that contain specific provisions or standards for the use of policy evaluation.

Figure 3.14. Guidelines containing standards for the use of policy evaluation



Note: n=31 (26 OECD member countries). 11 countries (9 OECD member countries) answered that they do not have guidelines to support the implementation of policy evaluation across government. Answers reflect responses to the question, "Do the guidelines contain specific guidance related to the: (Check all that apply)".
Source: OECD Survey on Policy Evaluation (2018).

Few countries have developed specific guidelines aimed at policy-makers for the uptake of evaluation evidence. These include New Zealand, Japan and Costa Rica (Box Box 3.24). In the US, the use of evaluation findings is an important part of the 2010 Federal Performance Framework (OMB, 2010^[178]). Overall, OECD data shows that the development of guidelines specifically dedicated to use of policy evaluations aimed at policy and decision-makers is a relatively recent practice, suggesting an increased awareness by countries of the need to development demand for evidence.

Box 3.24. Guidelines for the use of evaluation evidence

New Zealand: Making sense of evidence: A guide to using evidence in policy (2018)

New Zealand's Social Policy Evaluation and Research Unit (Superu) released this guide to provide central and local governments, the voluntary sector and the community with a structured approach to using evidence in every stage of the policy development cycle. This guide gives practical advice on:

- understanding the different sources and types of evidence, and the questions each is best suited to answer at each stage of policy development
- choosing and using evidence effectively according to three guiding principles: making sure that the evidence is appropriate, credible and transparent, and then explicitly stating how it is considered in every stage of the policy process.
- dealing with gaps in the evidence base and weak, uncertain and conflicting evidence
- taking into account different cultural values, and following a framework for bridging cultural perspectives, which is especially relevant in multicultural settings
- finally, getting stakeholders to engage early and commit to using evidence in policymaking through communicating findings (Superu, 2018^[179]).

Costa Rica: Guide for the use of evaluations: Guidelines for its implementation and follow-up on recommendations (2018)

The guide published by the Costa Rican Ministry of Planning and Economic Policy (Mideplan) provides support to decision-makers and those who execute policy interventions in concretely applying evaluation recommendations to improve the management of the public intervention evaluated. Precisely:

- It first defines the different types of use (instrumental, conceptual, persuasive and political) and emphasises their importance.
- Secondly, it details each step required to operationalise such use: analysis of recommendations, elaboration of a plan of action, implementation of the plan and analysis of its incidence. These steps include formalising and communicating decisions, identifying the actors and activities they involve, and elaborating a results report (Mideplan, 2018^[180]).

Japan: Policy Evaluation Implementation Guidelines (2005)

These guidelines state the importance of reflecting the results of an evaluation in the policy evaluated. They recommend individual administrative organs to prepare and release an evaluation report and compile a budget request to ensure that results are incorporated in policy planning. They suggest holding ministerial discussions on the results when the fiscal budget is being compiled or when important policy decisions are made to strengthen cooperation between the evaluation unit and ministry in charge of developing policies. The Ministry of Internal Affairs and Communications and administrative organs should explicitly state evaluation recommendations (such as the suppression, scaling up or down, or specific targeting of policy) when releasing the evaluation results in budget requests.

Sources: (Superu, 2018^[179]), (Mideplan, 2018^[180]), (The Ministry of Internal Affairs and Communication, 2005^[78]).

Other initiatives used by countries to promote use of evaluations include self-assessment tools to assess the capacity of organisations to demand and apply research. A key first step in enabling organisations to increase their ability to identify and assess research and use it in decision making is to examine the existing organisational capacity to access, interpret and use research findings (Kothari et al., 2009^[181]). Colombia, for example, launched the first guideline to construct an evidence gap map for civil servants, academics and external organization interested in using the existence evidence by performing a more robust approach.

Box 3.25. Guidelines for the construction of Evidence Gaps Maps: A tool for decision making in Colombia

The Colombian ministry of planning (DNP) created Guidelines for the construction of evidence gap maps (MBEs) to strengthen evidence-based decision-making. MBEs systematize and synthesize the evaluation results in a clear way, giving decision-makers an easy and comprehensive access to them and ultimately reinforcing use.

These guidelines can be used by any national public entities and international organisations interested in improving their decision-making processes. They present the steps required for the construction of an evidence gap map (MBE), accompanied by concrete examples and recommendations. They also describe the human resources needed to build the team responsible for constructing the MBE as well as the optimal planning for it.

Source: Taken from Colombia Ministry of Planning (2019), *Guideline for the construction of Evidence Gaps Maps: a tool for decision making*, Ministry of Planning

Increasing demand for evaluation by promoting competencies

In some countries, mechanisms to promote demand for evaluations are developed in addition to those aimed at promoting their supply. In fact, supply of evaluative evidence is not a sufficient condition for use: demand from primary intended users also needs to be there. Both research and practice indicate that despite the extensive production, communication and dissemination of evaluation reports, the use of evidence by decision makers remains limited, and the commitment of top management to evaluation activities remains low (Olejniczak and Raimondo, 2016^[182]).

Specifically, evaluation users – policy makers, in particular – can also face challenges related to their lack of competence to analyse and interpret evidence (Results for America, 2017^[166]), meaning that they do not have the appropriate skills, knowledge, experience and abilities to use evaluation results (Stevahn et al., 2005^[147]) (American Evaluation Association, 2018^[148]) (Newman, Fisher and Shaxson, 2012^[183]).

According to the survey, mechanisms aimed specifically at increasing demand for evaluations are less frequent than mechanisms aimed at promoting supply. Nevertheless, country practices reveal a wide range and approaches aimed at developing competences for use. This includes practices such as training, aimed at senior civil servants or policy professionals, and mentoring initiatives.

Understanding skills and competencies for policy evaluation

The OECD and the Joint Research Centre of the European Commission developed a Mapping of the relevant skills and competencies for Evidence Informed Policy Making from countries experiences. The skillset is presented below in Box 3.26. (A detailed discussion of this mapping can be found in the OECD forthcoming report on *Building Capacity for Evidence Informed Policy Making*).

Box 3.26. The skillset for Evidence Informed Policy Making

This skill-set is defined as a collective skill-set for the improvement of public service in the future and not as a full list of skills that each public servant needs to master. This skillset does not apply to one scenario; instead, it is of a cross-cutting character and can be applied on multiple occasions. It includes elements like critical thinking, systems thinking, and engaging with stakeholders. The skillset is defined by six (6) clusters, as follows:

Understanding evidence informed policy making

- Policy makers will be able to understand the role of evidence and its place in the policy making cycle, as well as the challenges and opportunities which come with the use of evidence.

Obtaining evidence

- Policy makers with this skill will be able to gather existing evidence in their own policy area. They will also be able to discuss evidence gaps and commission high quality evidence to solve these gaps.

Interrogating and assessing evidence

- Policy makers will be able to assess the reliability and appropriateness of evidence. They will have an ability to interrogate evidence by critically assessing its quality and context, using a range of techniques to challenge assumptions and biases.

Using and applying evidence in the policy making

- Policy makers will understand their own policy context and recognise possible uses of evidence in the policy cycle. They will be proficient in using innovative techniques like behavioural insights, and foresight to support policy design and implementation.

Engaging with stakeholders in evidence informed policy making

- Policy makers with this skill will have strong engagement and communication skills, including ability to provide information to different types of audiences and to engage and inspire variety of stakeholders.

Evaluating the success of evidence informed policy making

- Policy makers with this skill will understand different evaluation approaches and tools, and that evaluation should be built in the policy cycle and should serve to inform and improve Evidence informed policy making.

Source: Adapted from Building Capacity for Evidence Informed Policy Making: Lessons from country experiences, OECD (2020).

Training for policy makers and civil servants

Training refers to an active preparation based on appropriate approaches and strong guidance for a specific attendance. OECD's work on how to engage public employees for a high performing civil service highlights the importance of learning and training in a modern civil service to enable civil servants to continually update their skills and capacity to innovate (OECD, 2016^[184]). There is, therefore, a strong justification for investment in learning and training, and there is also a strong call from employers and employees for the need to invest in skills and competency development.

The work by the OECD on *Building Capacity for Evidence Informed Policy Making* (OECD, 2020^[99]) suggests that training for Senior Civil Service leadership is aimed at increasing managers' *understanding* of evidence informed policy making and policy evaluation, enabling them to become champions for evidence use. Intensive skills training programmes aimed at policy makers may be more focused on *interrogating and assessing* evidence and on *using and applying* it in policy making.

Training for Senior Civil Service leadership can include training courses or seminars given by national schools of government in the context of their leadership programmes or specific training courses developed by ministries or agencies. In Canada, for example, the executive training in research application (EXTRA) programme provides support and development for leaders in using research. The programme is targeted towards leaders in the healthcare field. The programme's objectives are that after the completion of the training, participants will be able to use evidence in their decision-making and will be able to train their co-workers and bring about organizational change.

Intensive skills training programmes geared towards policy makers can provide them with the necessary skills to increase the use of evidence in their work. Through such trainings, policy makers not only learn new skill but often also have increased motivation to use evidence and many become research champions and train or mentor others (Haynes et al., 2018^[173]). Such trainings can take the form of workshops, masterclasses or seminars (see Box 3.27 for such examples).

Box 3.27. Intensive skills training and mentoring programmes

In the **UK**, the **Alliance for Useful Evidence** organises an evidence masterclass where policy makers can learn about how to use evidence in their policy work and can practice their new skills through simulations. Through this programme, policy makers are able to build their confidence in compiling, assimilating, distilling, interpreting and presenting evidence. Participants learn how to find research that is relevant to their policy question, and they develop their ability to assess the quality and trustworthiness of research.

Mexico has also implemented capacity-building initiatives concerning Regulatory Impact Assessment. Training seminars were held by Mexico's ministry of the economy, for Federal and Provincial officials on how to draft and implement Regulatory Impact Assessments (RIA). The learning programme provided a step-by-step methodology on how to produce and analyse impact assessments in practice using guidance, case studies and advice from peer government officials, experts and OECD insights (OECD, 2020^[99]).

South Africa has a longstanding history of initiatives to improve the demand side for evidence use in policy making, including the implementation of workshops and a mentorship programme throughout government. The programme was created to address the disconnect between the widespread support for EIPM in principle and its practical application. The workshops and group mentoring were geared towards laying the foundations for individuals to acquire evidence informed policymaking skills. The group orientation created an environment in which there was greater acceptance of the value and practice of EIPM and therefore made individual mentoring possible. Those individuals were then able to mentor their colleagues on integrating evidence into their work.

Sources: adapted from OECD (Forthcoming), Building capacity for evidence informed policy making: lessons from country experience.

Mentoring programmes

Mentoring initiatives, on the other hand, refer to more personalised guidance on ‘real-world’ applications. For this initiative, the expertise and interpersonal skills of the mentors are key for the credibility of the process. While evidence suggests that mentoring initiatives can be successful in helping policy makers in using and applying evidence in their work, these types of programmes are less frequently used. One exception includes the Data for Decision Making (DDM) programme that was implemented in Mexico, which mainly relied on mentoring to improve the use of evidence in health policy-making (Punton et al., 2016^[185]). Box 3.27 above provides details on South African initiatives regarding mentorship programmes for government.

Creating an evaluation market place by embedding use of evidence in the institutional set-up

While individual competencies are important, formal organisations and institutional mechanisms set-up a foundation for evidence-informed policy making that can withstand transitions between leadership (Results for America, 2017^[166]). The use of evaluations is intimately linked to organisational structures and systems, insofar as they create a fertile ground for supply and demand of evaluations to meet.

Institutional or organisational mechanisms which enable the creation of an evaluation market place can be found either at the level of specific institutions, such as management response mechanisms, or within the wider policy cycle, such as through the incorporation of policy evaluation findings into the budget cycle or discussions of findings at the highest political level.

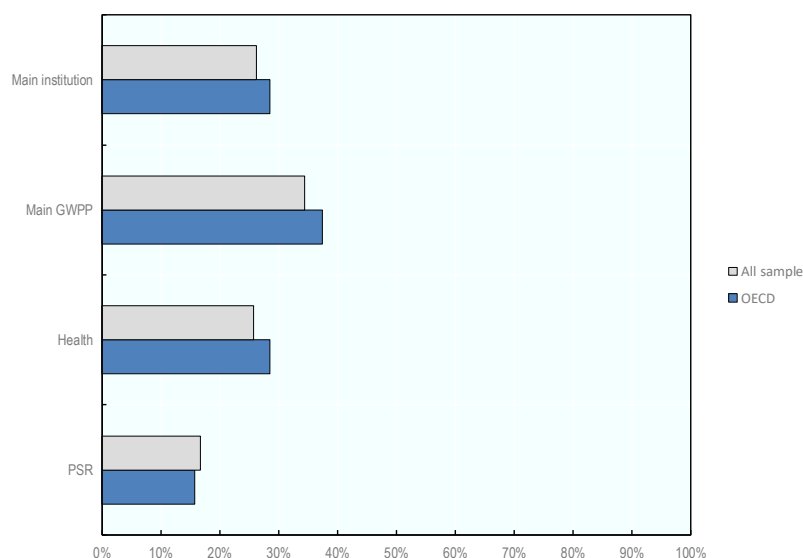
Management response mechanisms at the level of specific institutions

The first level of use for evaluations is management response mechanisms. Management response mechanisms indicate whether senior management partially agrees or disagrees with the assessment and strategic recommendations contained in a policy evaluation. The reason for agreement or disagreement is provided, and actions to be taken in response to the evaluation are described.

According to OECD data, the use of formal management response and follow-up systems is relatively infrequent. Amongst country respondents, 11 main institutions in charge of policy evaluations (10 in OECD countries), and a similar sample of Health Ministries (8) used management response mechanisms to react to internal or external evaluations. Some exceptions include management response mechanisms in Mexico and Costa Rica (see Box 3.28).*

The information from the OECD survey also suggests that a larger number of countries use management response and follow-up mechanisms for the evaluation of government-wide policy priorities. In Japan, for instance, the government submits each year a report to the Diet (Houses of Representatives and Councillors) on the status of policy evaluation and on how its results have been reflected in policy planning and development. In Korea, for instance, based on the review of evaluation results, improvements to be made to evaluated policies are identified and the evaluation plan for the subsequent year is adapted accordingly.

Figure 3.15. Management response mechanisms at the level of specific institutions



Note: For the main institution n=42 (35 OECD countries). For the main institution on government-wide policy priorities n=29 (24 OECD countries). 4 countries (all OECD) answered that they do not have government-wide policy priorities. 9 countries (7 OECD) answered that they do not evaluate their government-wide policy priorities. For the Health ministries n=31 (28 OECD). 9 countries (7 OECD) did not participate on this survey. 2 countries (1 OECD) are not included as they answered that none of the policies that fall in their institution's responsibility are evaluated. For the PSR ministries n=25 (20 OECD). 11 countries (10 OECD) did not participate in this survey. 6 countries (5 OECD) are not included as they answered that none of the policies that fall in their institution's responsibility. Answers reflect responses to the question "How does your government promote the use of the findings of policy evaluations?" and "How does your institution promote the use of the findings of policy evaluations? (Check all that apply)" focused on "A management response mechanism at the level of specific institutions is in place".

Source: OECD Survey on Policy Evaluation (2018).

Box 3.28. Management response mechanisms in Mexico and Costa Rica

Mexico implemented a mechanism to establish a follow-up process on external evaluation recommendations, which defines the actors responsible for constructing the tools that will track the aspects of programmes and policies to be improved. The Mexican National Council for the Evaluation of Social Development Policy (CONEVAL) gives a prize to federal ministries and states who contribute to the generation and use of evaluations results to improve policies, as well as to the development of their staff's skills for that purpose.

Costa Rica's ministry of national planning and political economy (Mideplan) developed a guide for the use of evaluations. It advises that an action plan should be developed based on an analysis of evaluations' recommendations. This requires convening key actors and stakeholders, and formulating and communicating decisions. The action plan should then be formalised by defining activities, roles and responsibilities, establishing expected results, and finally validating and communicating the plan. Implementing the plan requires incorporating it in instruments of organisational planning, monitoring compliance with its activities and generating a report. Finally, the impact of the action plan should be assessed by collecting and analysing the reports, consolidating its results, and publishing them.

Sources: OECD (2018) Survey on Policy Evaluation, (Mideplan, 2018^[180]).

Management response systems can also be informal. In the United Kingdom/Great Britain, for example, there are formally no specific requirements on how policy evaluation results are to be followed-up on. However, most institutions will develop some form of ministerial/management response to the results.

The role of knowledge brokers

Knowledge brokers are pivotal actors that connect knowledge producers and users in networks where knowledge and evidence is produced (Olejniczak and Raimondo, 2016^[182]). They can help to facilitate policymakers' access to the results of evaluations and to research evidence by helping them to navigate research material that may be unfamiliar. They can also help to articulate policymakers' needs, constraints and expectations, translating them for researchers who may be unfamiliar with the policy process (see Box 3.29).

Box 3.29. The role of knowledge brokers

Knowledge brokers play a key role in strengthening the relationship and collaboration between evidence producers and policymakers. A knowledge broker is either an individual, organisation, or structure that shares information, strengthens capacity and builds partnerships.

Governments can rely on knowledge brokers to improve their communication towards the evidence community regarding their particular needs and expectations for policymaking. On the other hand, knowledge brokers may also help evidence producers “translate” their results to policy makers, by synthesising them, disseminating them and expressing them in a clear and relevant manner

Overall, knowledge brokers have to both understand the technicalities of the research and evaluation world, as well as the practicalities of the actions and decisions taken by policy makers and the political, economic and social factors that influence them. More precisely, they can undertake the following activities to effectively transfer the knowledge they create to policy-makers and society more broadly:

- identifying the information gaps and needs of the users of evidence (decision-makers and policy actors)
- acquiring quality evidence from appropriate sources and in a timely manner (at the right stage of an intervention)
- transferring evidence to users by translating it in an appealing, tailored and actionable message, which may involve discussion and persuasion
- building networks between evidence producers and users to facilitate interactions and collaboration, ultimately allowing capacity building and dissemination
- accumulating evidence over time to build a robust and diversified evidence base, which requires building institutional capabilities for extracting useful evidence
- fostering an evidence-based culture by organising interactive workshops with decision makers to develop their skills and commitment to using evidence.

Source: (Results for America, 2017^[166]). (Olejniczak, Raimondo and Kupiec, 2016^[64]).

Knowledge brokers can take on a variety of forms, ranging from individual professionals (such as Government chief science advisors in some countries, or ministerial advisors) to dedicated organisations. In terms of institutions, some are specifically connected to knowledge producers, such as brokering units within academic institutions (for example, the Centre for Evaluation and Analysis of Public Policies in Poland and the Top Institute of Evidence-Based Education Research in the

Netherlands). Other approaches, on which this section will focus, locate the function closer to decision makers, either within government or at arms' length.

Evaluation units as knowledge brokers

Firstly, evaluation units or advisory bodies within ministries play an important knowledge brokerage role within their institution, as they convey their findings to departments responsible for planning and implementing interventions in their institution. As such, evaluation units are the first knowledge brokers, as they typically act as intermediaries between knowledge producers (evaluators) and actors involved in policy decisions (Olejniczak and Raimondo, 2016^[182]).

In France, for instance, many of the knowledge brokerage functions are integrated within the ministries. Analytical directorates in the ministries of labour (DARES), social affairs (DREES) or the environment (CGEDDE) provide strategic advice and access to evidence, and integrate the knowledge broker functions within the day-to-day work of the ministries.

Bodies at arm's length of government

Other countries have seen the development of knowledge brokerage organisations at arm's length of government. These units may function with a certain degree of independence, for instance in terms of staffing or budget, but receive government funding. These units often concentrate on one thematic area of specialisation. Examples of such organisations are the Australian Institute of Family Studies (AIFS), the Research and Evaluation Unit Department of Children and Youth Affairs in Ireland and the What Works Network in the United Kingdom (see the section on clearinghouses for a description of the What Works Network). Others, like Australia's Productivity Commission (see Box 3.30), are cross-disciplinary. The experience of the Productivity Commission is in many regards exemplary in terms of communication and external engagement. (For a full review of policy advisory bodies see (OECD, 2017^[186])).

Box 3.30. The experience of Australia's Productivity Commission in communication and public inquiries

The Australian Productivity Commission provides analysis and recommendations on specific policies and a range of economic, social and environmental issues. One of the main activities of the commission is the **communication** of its ideas and analyses, a key determinant for the use of evaluation results. First, the Commission's research reports are formally presented for discussion to the Australian Parliament through the Treasurer. Then, as the Commission is statutorily required to promote a public understanding of policy issues, it directs its reports and other activities at the wider community. For instance, all draft reports and preliminary findings are shared with the public for discussion through workshops, presentations and forums.

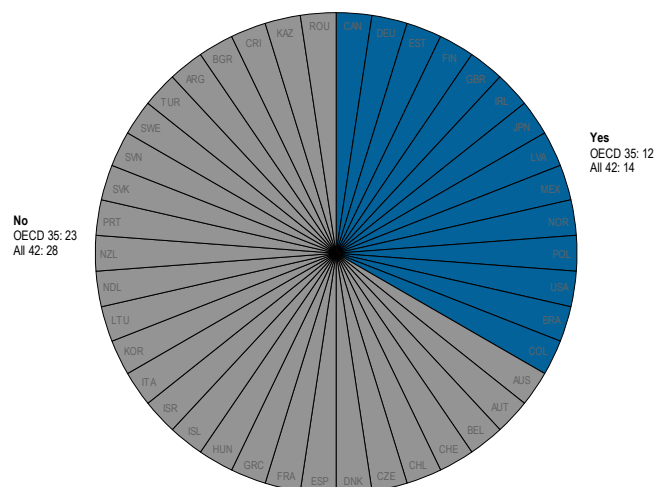
Public inquiries are another means used by the commission to handle policy issues that require significant public exposure and consultation. When policies have complex consequences or potential important impacts on society, citizens are consulted and their perspectives considered during policy formulation. To reach as many citizens and foster their active involvement, these public inquiries are widely advertised. For instance, the commission's inquiry on disability care and support received 1062 submissions, led to the visits of 119 organisations and included 23 days of public hearings.

Sources: Australian Government, "About the Commission" and "Core Functions". Accessed September 2nd 2019. <https://www.pc.gov.au/about>, <https://www.pc.gov.au/about/core-functions>

Coordination platforms or units across government

The results of the survey show that some governments (14 countries, of which 12 OECD countries) have established dedicated cross-governmental units to champion the use of policy evaluations in a horizontal manner (see Figure 3.16). These coordination platforms, or knowledge brokers, can take on a variety of organisational form, often close to the Centre of Government. These include the UK Cabinet Office's 'What Works' team, the evidence team within the Office of Management and Budget in the US (see Box 3.31) and France Stratégie, a think tank attached to the Prime Minister's Office in France.

Figure 3.16. Coordination platforms across government to promote use of evidence



Note: For the main institution n=42 (35 OECD member countries). Answers reflect responses to the question "How does your government promote the use of the findings of policy evaluations?" focused on "A coordination platform across government to promote the use of evidence (produced by policy evaluations) in policy making".

Source: OECD Survey on Policy Evaluation (2018).

Box 3.31. The role of the US Office of Management and Budget in promoting the use of evaluation

The United States' Office of Management and Budget (OMB) has a dedicated evidence team that acts as a central hub of expertise across the federal government. The team works with other OMB offices in order to set research priorities and ensure the use of appropriate evaluation methodologies in federal evaluations. As of July 2019, the team has created an interagency council that regroups Evaluation Officers. This council is intended to serve as a forum for officers to exchange information and advise the OMB on issues affecting the evaluation functions such as evaluator competencies, best practices for programme evaluation and evaluation capacity building. The council also allows for coordination and collaboration between evaluators and the government. It plays a leadership role for the larger Federal evaluation community. To ensure that evidence is used in policy design, the Evidence Team is also actively involved offering technical assistance to Federal Agencies.

Source: Clark, C. (2019) "OMB Moving Ahead to Steer Agencies on Evidence-Based Policymaking" <https://www.govexec.com/management/2019/07/omb-moving-ahead-steer-agencies-evidence-based-policymaking/158381/>

Of the 13 countries making use of such platforms, only five have mandates focused on matching supply and demand for evaluations (see Table 3.7).

Table 3.7. Mandates of coordination platforms

	Mapping the evidence brokerage function across government as way to foster systematic use of evidence	Ensure that policy evaluation and resources for evidence use are directed to inform policy design for government priorities	Ensure that the production of evidence matches the demand of policy makers in terms of timing and format	Enable the sharing of policy evaluations and of evidence results to practitioners and local governments to improve service delivery	Facilitate international cooperation in evidence production and use to enable efficiency gains
Canada	○	○	○	○	○
Estonia	●	●	○	○	○
Finland	○	●	●	●	○
Germany	○	●	●	●	●
Great Britain	○	●	●	●	○
Ireland	○	●	●	●	○
Japan	○	○	○	○	○
Latvia	●	●	○	●	○
Mexico	○	●	●	●	●
Norway	○	○	○	●	○
Poland	●	●	○	●	●
United States	●	○	○	●	○
Brazil	○	●	○	○	○
Colombia	○	●	○	●	●

Note: n=14 (12 OECD). Answers reflect responses to the question "What functions are being carried out by this coordination platform". The information reported here refers only to the countries that selected the option: "A coordination platform across government to promote the use of evidence (produced by policy evaluations)." in the question "How does your government promote the use of the findings of policy evaluations". In the option "other", Poland reported "assessment of evaluation reports influencing on robustness of evaluations", and Japan "information sharing".

Source: OECD Survey on Policy Evaluation (2018)

A majority of countries have mandates relating to ensuring that evaluation resources are directed to inform policy-design and decision-making. The following countries report that the coordination platform plays a role in mapping the evidence brokerage function: Estonia, Latvia, Poland, United States, , while Germany, Mexico, Colombia and Poland are the only countries that have explicitly attributed a role in facilitating international cooperation for use of evaluations to the platforms.

For instance, Japan's the Ministry of Internal Affairs and Communications (MIC) prepares an annual report on the status of policy evaluations carried out by the ministries and how the results of the evaluations have been reflected in policy planning and the development process. The MIC then aggregates the results of the evaluations conducted by ministries on the 'Portal Site for Policy Evaluation'.

Embedding the use of evaluation findings into policy planning/making processes

Incorporation of evaluation findings in the budgetary cycle

Incorporation of evaluation findings in the budgetary cycle is one of the most commonly used mechanism for the promotion of use of evaluations (see Figure 3.17). In fact, the results of the survey show that half of surveyed countries report that they incorporate evaluation evidence into the budgetary cycle. Sectoral respondents seem to incorporate such evidence less, with 35% (11 countries) of health respondents using evaluation evidence in budgetary decision-making and only 25% (6 countries) of public sector reform respondents.

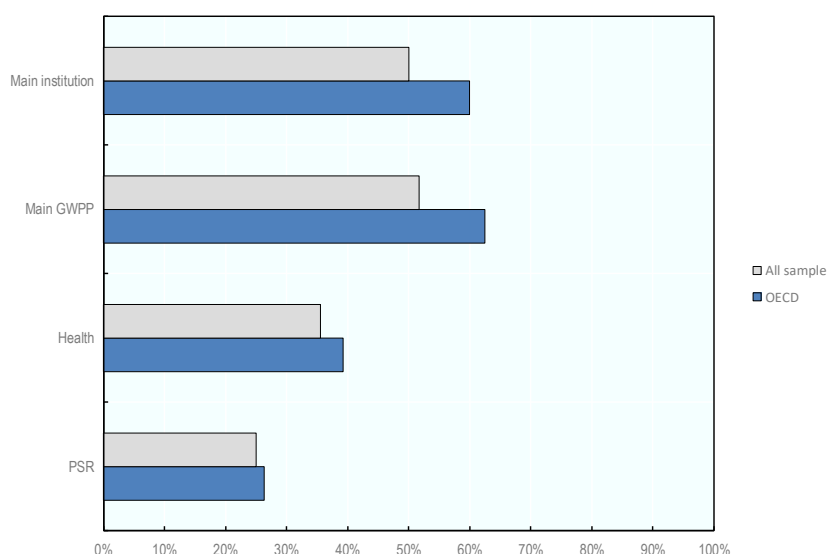
Country practices in this regard can take on a variety of forms depending on:

- the nature of the evidence produced: spending reviews or policy evaluation.
- the extent to which this evidence will impact budgetary decisions.

According to data from the budgeting and public expenditures survey (2018) (OECD, 2019^[12]), spending reviews are a widely used tool in OECD countries as part of the budget cycle. As discussed in chapter Chapter 1. , spending reviews produce performance evidence on programmes and policies. Nevertheless, while spending reviews focus on the effectiveness and efficiency of currently funded programmes in order to propose options for savings and fund reallocations, policy evaluations also look at impact of public interventions. Spending reviews also need to be informed by evaluations and an assessment of the effectiveness of programmes (Robinson, 2014^[187]) (Smismans, 2015^[14]) (The World Bank, 2018^[15]).

Many OECD countries (27 out of 33 respondents (OECD, 2019^[12]) make use of spending reviews and the evidence they produce in their budgetary cycle. Some countries, such as Denmark, the Netherlands or Germany, conduct *ad hoc* spending reviews to inform some allocation decisions every year. Others have used more comprehensive spending reviews, typically on a rolling basis over the period of an electoral mandate. The Irish Department of Public Expenditure and Reform has introduced a rolling spending review process, which consists of examining national expenditures over a three-year period and assessing the effectiveness of existing programmes. The three-year rolling nature of the review enables building up expertise and awareness of the process, and allows analysts to revisit emerging issues and further embed an evaluation culture across the Public Service.

Figure 3.17. Incorporation of policy evaluation findings into the budget cycle

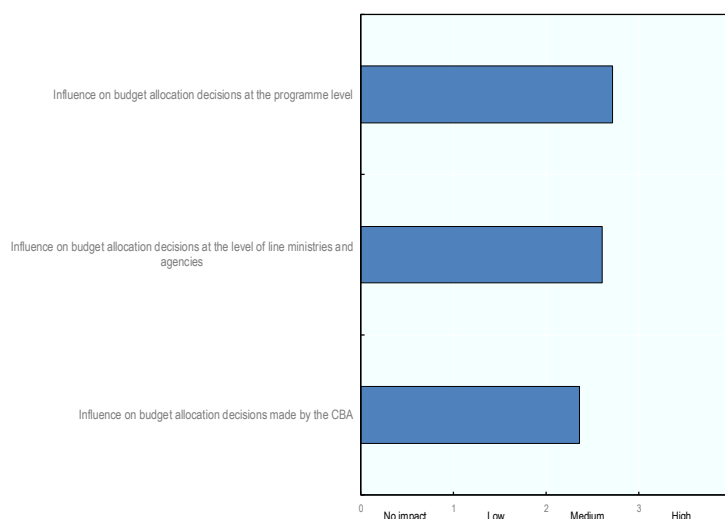


Note: For the main institution n=42 (35 OECD). For the main institution on government-wide policy priorities n=29 (24 OECD). 4 countries (all OECD) answered that they do not have government-wide policy priorities. 9 countries (7 OECD) answered that they do not evaluate their government-wide policy priorities. For the Health ministries n=31 (28 OECD). 9 countries (7 OECD) did not participate on this survey. 2 countries (1 OECD) are not included as they answered that none of the policies that fall in their institution's responsibility are evaluated. For the PSR ministries n=25 (20 OECD). 11 countries (10 OECD) did not participate on this survey. Moreover, 6 countries (5 OECD) are not included as they answered that none of the policies that fall in their institution's responsibility are evaluated. Answers reflect responses to the question "How does your government promote the use of the findings of policy evaluations" and "How does your institution promote the use of the findings of policy evaluations? (Check all that apply)" focused on "Incorporation of policy evaluation findings into the budget cycle".

Source: OECD Survey on Policy Evaluation (2018).

According to the results of the survey on performance budgeting for a sample of 20 countries, the impact of policy evaluations on budget decisions remains relatively limited compared to spending reviews.

Figure 3.18. Influence of evaluation findings on budget allocation decisions



Note: All OECD countries, n=28. 7 countries did not provide answers to this question. Answers reflect responses to the question, "At what level, and to what extent, do evaluation findings influence budget allocation decisions?".

Source: OECD Survey on Performance Budgeting (2018).

One exception includes the budgetary cycle in Lithuania, where the office of Government, together with the ministry of finance, summarises the results of evaluations in preparation for budget negotiations, in a note that provides information on progress achieved by the agency since the evaluation and any implementation gaps. Another such exception is the budgetary cycle in Canada (see Box 3.32).

Box 3.32. Use of evaluation findings in the budgetary cycle in Canada

The Treasury Board's reviews, expenditure and allocation decisions are required to be informed by evaluation findings. Such use of evaluation findings is ensured by requiring organisations in charge of policymaking to have to seek permission from the Treasury Board to obtain expenditure authority for policies, programmes and projects. These organisations have to address a list of practical questions related to evaluation, which can be found on the Canadian Government's website, when drafting a submission for the Treasury Board.

The organisation making a policy proposal has to define the expected results of the policy in light of existing policies, state whether an evaluation has been conducted and if so, share its results and otherwise state whether a future evaluation is planned. The policy proposal has to indicate whether the head of evaluation or head of performance measurement has been consulted in the development of the policy, which in any case has to be supported by relevant evidence, and eventual unfavourable evidence has to be discussed.

Source: OECD (2018) Survey on Policy Evaluation.

Finally, evaluative evidence may be used in a more or less systematic manner in the budget cycle. For instance, the OECD has identified four main models of performance budgeting, which reflect the different strength of the links between performance evidence and budgeting (OECD, 2019^[12]):

- presentational (evidence presented separately from the main budget document)
- performance informed (performance evidence included within the budget document that is presented on the basis of programmes)
- managerial
- direct performance budgeting (direct link between results and resources).

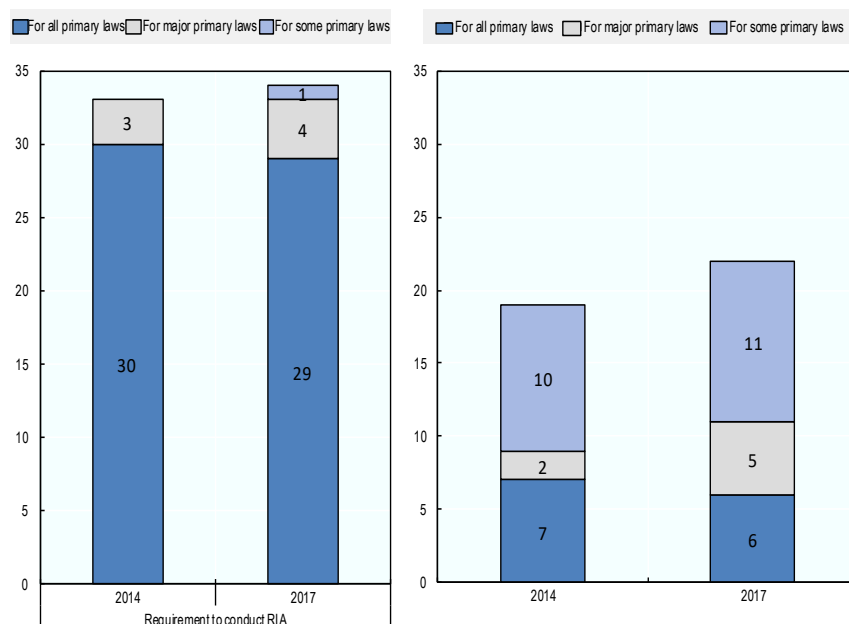
In most OECD countries, performance evidence is included in the budget cycle according to one of the first three approaches. France offers an example of strong links between key performance indicators established at the national level, their *ex post* evaluation and the budget cycle. France's organic budget law (*loi organique relative aux lois de finances* (LOLF)) groups expenditures by "missions" and programmes, which are each associated with policy objectives and performance indicators. Each budgetary cycle sees programmes from the previous cycles' budgetary being evaluated against these objectives and indicators in annual performance reports (*rapport annuels de performance*). These evaluations are included in the annex of the main budget document, which is examined by Parliament.

Another long standing example of embedding evaluations in policy making is the domain of regulatory policy, where there are requirements both for using evaluation *ex ante* as part of the RIA process and requirements related to the implementation and use of evaluations in laws and policies. An increasing number of laws and regulatory acts contain clauses that include formal requirements for policy evaluation. The OECD Council Recommendation on Regulatory Policy and Governance makes numerous references to evaluation as a part of promoting evidence-based decision making through including *ex ante* and *ex post* assessment of regulations (OECD, 2012). The OECD is currently developing best practice principles for RIA (GOV/RPC(2018)12/REV2) and also for *ex post* review as the attention to regulatory quality is increasingly shifting the focus not only on the evaluation of the impact *ex ante*, but also to the evaluation *ex post* (GOV/RPC(2018)5/REV2).

In the case of *ex ante* evaluations, these requirements can promote the use of evaluations as the evidence produced through the assessment is *in fine* meant to inform decision makers on whether and how to regulate to achieve public policy goals (OECD, 2018^[172]). Regulatory impact assessment (RIA) is “a systematic process of identification and quantification of benefits and costs likely to flow from regulatory or non-regulatory options for a policy under consideration. A RIA may be based on benefit-cost analysis, cost-effectiveness analysis, business impact analysis, etc.” (OECD, 2018^[172]). RIA can be an important tool for promoting evidence-information policy-making agenda.

RIA is now required in almost all OECD countries for the development of at least some regulations and implementation gaps are slowly diminishing (OECD, 2018^[172]).

Figure 3.19. Formal requirements to conduct RIA and *ex post* evaluation of primary laws



Note: Data for OECD countries is based on the 34 countries that were OECD members in 2014 and the European Union. Data on new OECD member and accession countries 2017 includes Colombia, Costa Rica, Latvia and Lithuania.³

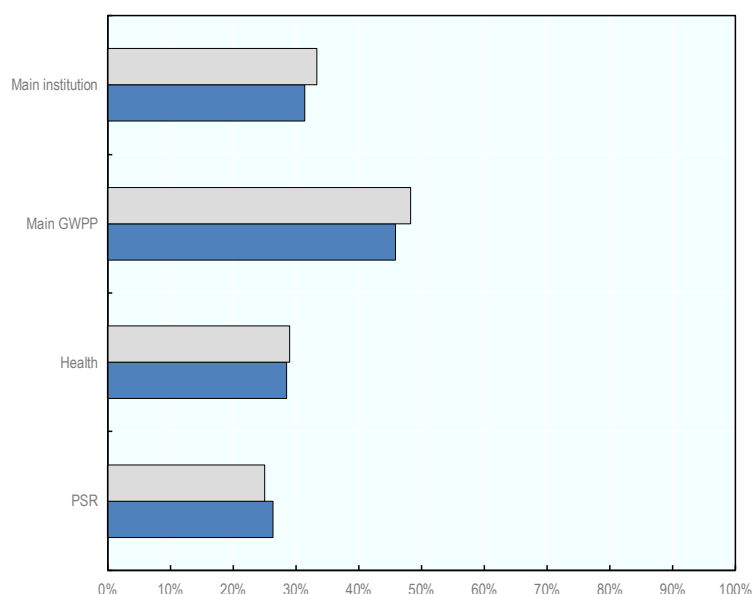
Source: Indicators of Regulatory Policy and Governance Surveys 2014 and 2017, <http://oe.cd/ireg>.

Ex-post evaluations of regulations seek to establish whether laws and regulations continue to be fit for purpose. At the same time, *ex post* evaluations provide an opportunity to assess whether there are better means of achieving the original policy goals and thereby further enhance societal welfare. According to data from the indicators of regulatory policy and governance survey (see (OECD, 2018^[123]), *ex post* review of regulations remains less institutionalised than *ex ante* assessments, with fewer countries having formalised arrangements. Less than one third of OECD countries have systematic requirements for *ex post* evaluation of regulations (OECD, 2018^[172]).

Using evaluation at the Centre of Government, and monitoring of government wide policy priorities

Countries have also used various mechanisms to discuss evaluation results at the highest political level. This practice is more frequent for the evaluation of government-wide policy priorities (GWPP), with about half of countries (14, of which 11 OECD countries) discussing evaluation findings at the level of the Council of Ministers (or equivalent), compared to a third of countries (14, of which 11 OECD countries) for policy evaluations in general. In Korea, for instance, in the context of the “100 Policy Tasks” five-year plan, evaluation results are discussed at the council of ministers (See Box 3.33).

Figure 3.20. Discussion of evaluation findings at the Centre of government



Note: For the main institution n=42 (35 OECD member countries). For the main institution on government-wide policy priorities n=29 (24 OECD). 4 countries (all OECD) answered that they do not have government-wide policy priorities. 9 countries (7 OECD) answered that they do not evaluate their government-wide policy priorities. For the Health ministries n=31 (28 OECD). 9 countries (7 OECD) did not participate on this survey. 2 countries (1 OECD member country) are not included as they answered that none of the policies that fall in their institution's responsibility are evaluated. For the PSR ministries n=25 (20 OECD member countries). 11 countries (10 OECD) did not participate on this survey. 6 countries (5 OECD) are not included as they answered that none of the policies that fall in their institution's responsibility are evaluated. Answers reflect responses to the question “How does your government promote the use of the findings of policy evaluations” and “How does your institution promote the use of the findings of policy evaluations? (Check all that apply)” focused on “Discussions of evaluation findings at the Council of Ministers (or equivalent)”.

Source: OECD Survey on Policy Evaluation (2018).

Box 3.33. Discussion of evaluation results at the Council of Ministers in Korea

In Korea, according to the Government performance evaluation implementation plan, evaluation results have to be discussed at the council of ministers. This plan is part of a larger five-year plan that requires ministerial and vice-ministerial agencies to carry out 100 policy tasks to turn the Republic into a more people-centred democracy. Ministerial capacity is evaluated with regards to these 100 policy tasks on the basis of job creation, attainment of targets and policy impact. Ministries are encouraged to take continued interest in this evaluation plan and give inputs on the implementation of the evaluated tasks by participating in forums within the state administration to discuss evaluation results. According to these results, the evaluation plan for the subsequent year is adapted and rewards are given to the best performing agencies, incentivising them further to make evidence-based decisions. Lastly, when necessary, the Prime minister presides over Government performance evaluation committee meetings, during which evaluation reports are reviewed.

Source: OECD (2018) Survey on Policy Evaluation

Other countries have set-up specific committees or councils, most often at the centre of government, in order to follow-up on the implementation of policy evaluations and/or discuss their findings. The Brazilian committee for monitoring and evaluation of federal public policies is an example of such committee, which brings together high-level representatives from the executive (Presidency of the Republic, ministry of finance, ministry of planning, and the ministry of transparency) and from the comptroller general of the Union (CGU) (see Box 3.34).

Box 3.34. The Brazilian Committee for Monitoring and Evaluation of Federal Public Policies

In 2016, the Brazilian Government established the committee for monitoring and evaluation of federal public policies (CMAP) with the objective of encouraging the use of evaluation results to improve public policy outcomes and performance, the allocation of resources, and the quality of public spending. The Committee involves the following institutions: ministry of planning, development and management, ministry of finance, ministry of transparency, the Union's general comptroller, and the civil house of the Presidency. They meet periodically to monitor and evaluate the public policies selected by the CMAP and accordingly propose alternative designs and adjustments to them. All policymakers in charge of the evaluated policies are invited to participate in the CMAP's evaluation activities. Moreover, although not always in a systematic way, most of evaluation findings are involved in broader political discussions on public policy. The CMAP has thereby been able to promote several reforms in the legal framework and design of evaluated policies. The CMAP can be seen as an effective mechanism for fostering evaluation use at the highest political level thanks to its composition, consisting of central ministries responsible for the public budget, public resources and political coordination.

Source: OECD (2018) Survey on Policy Evaluation.

At the sector level, such councils or committees can be set-up on an *ad hoc* basis in order to follow-up on the implementation of recommendations from policy evaluations. In Canada, for example, after the review of federally funded pan-Canadian health organizations was completed in March 2018, an implementation steering group (ISG) was formed to develop a detailed implementation plan and provide advice to health Canada on how to move forward with the recommendations of the review.

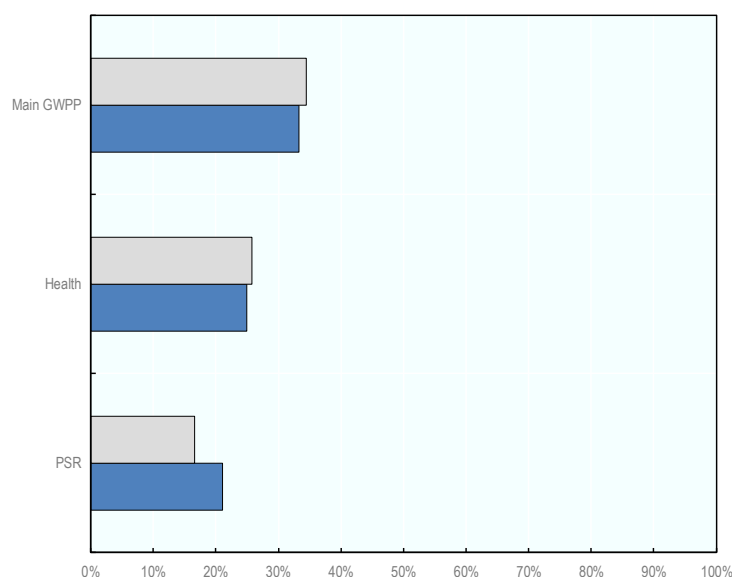
The role of institutions beyond of the executive

The role of Parliaments

Beyond their role as evaluation producers, parliaments have a particular role to play in promoting the use of evaluations. First, as they contribute to ensure accountability, parliaments have played an important role in promoting a more structured or systematic approach to conducting evaluations (Gaarder and Briceño, 2010^[55]). For instance, parliaments have been instrumental in increasing evaluation use by: promoting the use of evaluative evidence in the budgetary cycle by requiring more performance data on government spending, introducing evaluation clauses into laws and commissioning evaluations at the committee level in the context of hearings (Jacob, Speer and Furubo, 2015^[63]). Second, parliaments rely on verifiable and sound data on which they can base their policy initiatives and can thus push for the establishment of a structured approach to gather this information. Most parliaments have research and information services that help members of parliament order or request evaluation reports. Some even conduct evidence syntheses, thus playing a knowledge brokerage function (Jacob, Speer and Furubo, 2015^[63]). Independent fiscal institutions and parliamentary budget offices attached to parliament are some of the main users of such data. Parliaments are also recipients of evaluations conducted by other institutions or bodies, such as supreme audit institutions. In Denmark, the evaluations of *Rigsrevisionen* are handed over to parliament for a formal reaction. In Japan, the government submits each year a report to the diet on policy evaluation and on how the results of such evaluation have been reflected in policy planning and development.

Nevertheless, the results of the survey suggest that discussion of evaluation findings in parliament is a relatively infrequent practice. Only 10 countries evaluate findings related to the evaluation of their government-wide policy priorities in parliament; only 8 countries do so for health and 4 countries for PSR. France has recently promoted the “Spring of Evaluation” in 2018 as part of its parliamentary finance committee. This provides a platform to discuss work on policy evaluation, following the theme for evaluation and the planning that was agreed by the committee at the beginning of the year. Ministries are invited for hearings and invited to discuss the performance of the public policies for which they have responsibility. Three days of full discussions are then organised in a public hearing, which includes questions, discussions and the adoption of parliamentary resolutions.⁴

Figure 3.21. Discussion of evaluation findings in Parliament



Note: For the main institution on government-wide policy priorities $n=29$ (24 OECD). 4 countries (all OECD) answered that they do not have government-wide policy priorities. 9 countries (7 OECD) answered that they do not evaluate their government-wide policy priorities. For the Health ministries $n=31$ (28 OECD). 9 countries (7 OECD) did not participate on this survey. 2 countries (1 OECD) are not included as they answered that none of the policies that fall in their institution's responsibility are evaluated. For the PSR ministries $n=25$ (20 OECD). 11 countries (10 OECD) did not participate in the survey. 6 countries (5 OECD) are not included as they answered that none of the policies that fall in their institution's responsibility are evaluated. Answers reflect responses to the question "How does your government promote the use of the findings of policy evaluations?" and "How does your institution promote the use of the findings of policy evaluations? (Check all that apply)" focused on "Through discussion of evaluation findings in Parliament (or equivalent)".

Source: OECD Survey on Policy Evaluation (2018).

The role of Supreme Audit Institutions in promoting the use

SAIs play a role in promoting the use of evaluations in three main ways. First, as part of their mandate, many SAIs assess the mechanisms through which governments manage performance evidence, which includes looking at how evidence is used in the budgeting process and others systems for managing information. Second, SAIs contribute to evaluation use by disclosing the results of the evaluations they conduct. (OECD, 2016^[21]).

Indeed, an INTOSAI survey shows that among 14 SAIs, a majority (62%) generally publishes their evaluations, of which 75% states that they are frequently covered by the media (INTOSAI Working Group on Evaluation of Public Policies and Programs, 2019^[188]). Supreme audit institutions may contribute to use of evaluations by promoting public awareness of their results. Other SAIs use active communication strategies to ensure the use of the evaluations they conduct. The Swiss federal audit office, for instance, uses advisory groups as "multiplier agents" to foster the use of its evaluations by helping them disseminate and communicate their results (Swiss Federal Audit Office, 2019^[189]).

Third, some SAIs contribute to use by assessing government entities' use of evidence in decision-making as part of their mandate to evaluate for results. For example, the US Government accountability office produces reports and recommendations targeted to both the executive and to Congress on the implementation of the US Government performance management modernization Act (GRPAMA), which gives the Office of Management and Budget (OMB) an important role in disseminating and integrating a results and performance based approach to public administration.

Notes

¹ See A data-driven public sector: Enabling the strategic use of data for productive, inclusive and trustworthy governance (Working paper) [GOV/PGC(2019)57].

² Eleven evaluations from: the National Agency of Evaluations-ANE, Ministry of Health, and academia, and their terms of references)

³ The statistical data for Israel are supplied by and under the responsibility of the relevant Israeli authorities. The use of such data by the OECD is without prejudice to the status of the Golan Heights, East Jerusalem and Israeli settlements in the West Bank under the terms of international law.

⁴ See the French Parliament website: www.ausimplementationconference.net.au/

References

- 115th Congress (2019), *Public Law No: 115-435 (01/14/2019) - Foundations for Evidence-Based Policymaking Act of 2018*, <https://www.congress.gov/bill/115th-congress/house-bill/4174>. [7]
- Acquah, D., K. Lisek and S. Jacobzone (2019), “The Role of Evidence Informed Policy Making in Delivering on Performance: Social Investment in New Zealand”, *OECD Journal on Budgeting*, Vol. 19/1, <https://dx.doi.org/10.1787/74fa8447-en>. [18]
- AEVAL (2015), *Practical guide for the design and implementation of public policy evaluations*(*Guía práctica para el diseño y la realización de evaluaciones de políticas públicas*), http://www.aeval.es/export/sites/aeval/comun/pdf/evaluaciones/Guia_Evaluaciones_AEVAL.pdf (accessed on 21 August 2019). [131]
- Alkin, M. and S. Taut (2002), “Unbundling evaluation use”, *Studies in Educational Evaluation*, Vol. 29/1, pp. 1-12, [http://dx.doi.org/10.1016/S0191-491X\(03\)90001-0](http://dx.doi.org/10.1016/S0191-491X(03)90001-0). [161]
- American Evaluation Association (2018), *Guiding Principles*. [148]
- American Evaluation Association (2015), *Core Evaluator Competencies*, <http://www.eval.org>. [151]
- Barnett, C. and L. Camfield (2016), “Ethics in evaluation”, *ournal of Development Effectiveness*. [117]
- Better evaluation (2019), *Review evaluation (do meta-evaluation)*, https://www.betterevaluation.org/en/rainbow_framework/manage/review_evaluation_do_meta_evaluation (accessed on 19 August 2019). [142]
- Bossuyt, J., L. Shaxson and A. Datta (2014), “Study on the uptake of learning from EuropeAid’s strategic evaluations into development policy and practice”, Evaluation Unit of the Directorate General for Development and Cooperation-EuropeAid (European Commission). [67]
- Bridgeland, J. and P. Orszag (2013), *Can Government Play Moneyball? - The Atlantic*, <https://www.theatlantic.com/magazine/archive/2013/07/can-government-play-moneyball/309389/> (accessed on 6 December 2018). [165]

- Brown, L. and S. Osborne (2013), "Risk and Innovation", *Public Management Review*, Vol. 15/2, pp. 186-208, <http://dx.doi.org/10.1080/14719037.2012.707681>. [60]
- Brown, R. and D. Newman (1992), "Ethical Principles and Evaluation Standards: Do They Match?", *Evaluation Review*, Vol. 16/6, pp. 650-663, <http://dx.doi.org/10.1177/0193841X9201600605>. [119]
- Bundesministerium für Finanzen and Bundesministerin für Frauen und öffentlichen Dienst (2013), *Handbuch Wirkungsorientierte Folgenabschätzung Arbeitsunterlage*, <http://www.oeffentlicherdienst.gv.at> (accessed on 28 August 2019). [152]
- Campbell, S. and G. Harper (2012), *Quality in policy impact evaluation: understanding the effects of policy from other influences (supplementary Magenta Book guidance)*, HM Treasury, <http://www.nationalarchives.gov.uk/doc/open-> (accessed on 9 July 2019). [129]
- Canada Treasury Board (2016), *Policy on results*, <https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=31300> (accessed on 20 September 2019). [8]
- Caroline Heider (2018), *The Three Pillars of a Working Evaluation Function: IEG's Experience*, <http://ieg.worldbankgroup.org/blog/three-pillars-working-evaluation-function-iegs-experience> (accessed on 22 August 2019). [133]
- Cinar, E., P. Trott and C. Simms (2018), "Public Management Review A systematic review of barriers to public sector innovation process", <http://dx.doi.org/10.1080/14719037.2018.1473477>. [65]
- Commission on Evidence-Based Policymaking (2017), *The Promise of Evidence-Based Policymaking: Report of the Commission on Evidence-Based Policymaking*, <https://www.cep.gov/report/cep-final-report.pdf> (accessed on 6 August 2019). [31]
- CONEVAL (2007), *Lineamientos generales para la evaluación de los Programas Federales de la Administración Pública Federal*, https://www.coneval.org.mx/rw/resource/coneval/eval_mon/361.pdf (accessed on 18 June 2019). [23]
- Cooksy, L. and M. Mark (2012), "Influences on evaluation quality", *American Journal of Evaluation*, Vol. 33/1, pp. 79-84, <http://dx.doi.org/10.1177/1098214011426470>. [159]
- Crowley, D. et al. (2018), "Standards of Evidence for Conducting and Reporting Economic Evaluations in Prevention Science", *Prevention Science*, Vol. 19/3, pp. 366-390, <http://dx.doi.org/10.1007/s1121-017-0858-1>. [45]
- Damschroder, L. et al. (2009), "Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science", *Implementation Science*, Vol. 4/1, p. 50, <http://dx.doi.org/10.1186/1748-5908-4-50>. [109]
- Departamento Nacional de Planeación (2016), *¿Qué es una Evaluación?*, <https://sinergia.dnp.gov.co/Paginas/Internas/Evaluaciones/%C2%BFQu%C3%A9-es-Evaluaciones.aspx>. [25]
- Department of the Prime Minister and Cabinet (2014), *Guide to Implementation Planning*, Australian government, <https://www.pmc.gov.au/sites/default/files/files/guide-to-implementation-planning.pdf> (accessed on 12 July 2019). [42]

- Department of the Prime Minister and Cabinet (Australia) (2013), *Policy Implementation* |, [85]
<https://www.pmc.gov.au/government/policy-implementation> (accessed on
 20 September 2019).
- Desautels, G. and S. Jacob (2012), “The ethical sensitivity of evaluators: A qualitative study
 using a vignette design”, *Evaluation*, Vol. 18/4, pp. 437-450, [116]
<http://dx.doi.org/10.1177/1356389012461192>.
- Dobbins, M. et al. (2009), “A randomized controlled trial evaluating the impact of knowledge
 translation and exchange strategies”, *Implementation Science*, Vol. 4/1, p. 61, [176]
<http://dx.doi.org/10.1186/1748-5908-4-61>.
- Estonian National Audit Office (2011), *The state of affairs with the legislative impact assessment*. [157]
- European Commission (2017), *Better Regulation Guidelines*, [http://europa.eu/about-eu/basic-
 information/decision-making/treaties/index_en.htm](http://europa.eu/about-eu/basic-information/decision-making/treaties/index_en.htm). [171]
- European Court of Auditors (2013), *Audit Guidelines on Evaluation*, [158]
[https://www.eca.europa.eu/Lists/ECADocuments/GUIDELINES_EVALUATION/Evaluation-
 Guideline-EN-Oct2013.pdf](https://www.eca.europa.eu/Lists/ECADocuments/GUIDELINES_EVALUATION/Evaluation-Guideline-EN-Oct2013.pdf) (accessed on 23 August 2019).
- European Environment Agency (2017), “EEA guidance document-policy evaluation”, [27]
<https://www.researchgate.net/publication/317594615>.
- Flay, B. et al. (2005), “Standards of Evidence: Criteria for Efficacy, Effectiveness and
 Dissemination”, *Prevention Science*, Vol. 6/3, pp. 151-175, [http://dx.doi.org/10.1007/s11121-
 005-5553-y](http://dx.doi.org/10.1007/s11121-005-5553-y). [50]
- Fleischer, D. and C. Christie (2009), “Evaluation use: Results from a survey of U.S. American
 evaluation Association members”, *American Journal of Evaluation*, Vol. 30/2, pp. 158-175, [162]
<http://dx.doi.org/10.1177/1098214008331009>.
- Flemig, S., S. Osborne and T. Kinder (2016), “Risky business—reconceptualizing risk and
 innovation in public services”, *Public Money & Management*, Vol. 36/6, pp. 425-432, [61]
<http://dx.doi.org/10.1080/09540962.2016.1206751>.
- France Stratégie (2016), *How to evaluate the impact of public policies: a guide for the use of
 decision makers and practitioners (Comment évaluer l'impact des politiques publiques : un
 guide à l'usage des décideurs et des praticiens)*, [128]
[https://www.strategie.gouv.fr/sites/strategie.gouv.fr/files/atoms/files/guide_methodologique_2
 0160906web.pdf](https://www.strategie.gouv.fr/sites/strategie.gouv.fr/files/atoms/files/guide_methodologique_20160906web.pdf) (accessed on 21 August 2019).
- France Stratégie, R. Desplat and M. Ferracci (2016), *Comment évaluer l'impact des politiques
 publiques ? Un guide à l'usage des décideurs et praticiens*. [130]
- France Stratégie, R. Desplat and M. Ferracci (n.d.), *Comment évaluer l'impact des politiques
 publiques ? Un guide à l'usage des décideurs et praticiens*. [134]
- Gaarder, M. and B. Briceño (2010), “Institutionalisation of government evaluation: balancing
 trade-offs”, *Journal of Development Effectiveness*, Vol. 2/3, pp. 289-309, [55]
<http://dx.doi.org/10.1080/19439342.2010.505027>.

- Gaspar, D. (2018), "Policy Evaluation: From Managerialism and Econocracy to a Governance Perspective", in *International Development Governance*, Routledge, <http://dx.doi.org/10.4324/9781315092577-37>. [28]
- Gaspar, D. (2018), "Policy Evaluation: From Managerialism and Econocracy to a Governance Perspective", in *International Development Governance*, Routledge, <http://dx.doi.org/10.4324/9781315092577-37>. [30]
- Gauthier, B. (2015), "Some pointers concerning Evaluation Utilization". [168]
- Goldstein, C. et al. (2018), *Ethical issues in pragmatic randomized controlled trials: A review of the recent literature identifies gaps in ethical argumentation*, BioMed Central Ltd., <http://dx.doi.org/10.1186/s12910-018-0253-x>. [51]
- Government Policy Analysis Unit (2017), *Global Evidence Policy Units: Finland*, https://www.ksj-indonesia.org/file_upload/Evidence-Policy-Unit-in-Finland-the-Government-Po-14Jun2017163532.pdf. [90]
- Greenhalgh, T. et al. (2004), "Diffusion of Innovations in Service Organizations: Systematic Review and Recommendations", *The Milbank Quarterly*, Vol. 82/4, pp. 581-629, <http://dx.doi.org/10.1111/j.0887-378X.2004.00325.x>. [110]
- Haynes, A. et al. (2012), "Identifying Trustworthy Experts: How Do Policymakers Find and Assess Public Health Researchers Worth Consulting or Collaborating With?", *PLoS ONE*, Vol. 7/3, p. e32665, <http://dx.doi.org/10.1371/journal.pone.0032665>. [170]
- Haynes, A. et al. (2018), "What can we learn from interventions that aim to increase policy-makers' capacity to use research? A realist scoping review", *Health Research Policy and Systems*, Vol. 16/1, p. 31, <http://dx.doi.org/10.1186/s12961-018-0277-1>. [173]
- Heider, C. (2017), *Rethinking Evaluation - Efficiency, Efficiency, Efficiency*, <https://ieg.worldbankgroup.org/blog/rethinking-evaluation-efficiency>. [29]
- Hildén, M. (2014), "Evaluation, assessment, and policy innovation: exploring the links in relation to emissions trading", *Environmental Politics*, Vol. 23/5, pp. 839-859, <http://dx.doi.org/10.1080/09644016.2014.924199>. [41]
- HM Treasury (2011), *The Magenta Book: Guidance for evaluation*, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/220542/magenta_book_combined.pdf (accessed on 18 June 2019). [9]
- HM Treasury (2011), *The Magenta Book: Guidance for evaluation*, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/220542/magenta_book_combined.pdf (accessed on 18 June 2019). [191]
- Howlett, M. (2019), "Policy analytical capacity and evidence-based policy-making: Lessons from Canada", https://doi.org/10.1111/j.1754-7121.2009.00070_1.x. [3]
- IGEES (2014), *Irish Government Economic and Evaluation Service*, <https://igees.gov.ie/> (accessed on 28 January 2019). [153]

- Impact Assessment Office (2018), “The Uncompleted Evaluation of Legislative in Italy: Critical Issues, Prospects and Good Practice”, <http://www.senato.it/service/PDF/PDFServer/BGT/01082854.pdf> (accessed on 23 September 2019). [94]
- Independent Evaluation Office of UNDP (2019), *UNDP Evaluation Guidelines*. [127]
- Innovate UK (2018), “Evaluation Framework, How we assess our impact on business and the economy”, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/681741/17.3253_Innovate_UK_Evaluation_Framework_RatherNiceDesign_V2_FINAL_WEB.pdf (accessed on 20 September 2019). [82]
- International Organisation of Supreme Audit Institutions (2019), *ISSAI 100: Fundamental Principles of Public-Sector Auditing*, <http://www.issai.orghttp://www.intosai.org> (accessed on 8 January 2020). [20]
- INTOSAI (2016), *Guidelines on the Evaluation of Public Policies*, <http://www.issai.org> (accessed on 22 August 2019). [155]
- INTOSAI (2010), *Program Evaluation for SAIs, A Primer*, <https://www.eurosai.org/handle404?exporturi=/export/sites/eurosai/.content/documents/materials/Program-Evaluation-for-SAIs.pdf> (accessed on 22 August 2019). [122]
- INTOSAI Working Group on Evaluation of Public Policies and Programs (2019), *Implementation of the INTOSAI GOV 9400 Guidelines: Survey Results*. [188]
- INTOSAI and Cour des Comptes (France) (2019), *Implementation of the INTOSAI GOV 9400 Guidelines: Survey results*. [97]
- Jacob, S. and Y. Boisvert (2010), *To Be or Not to Be a Profession: Pros, Cons and Challenges for Evaluation*, <http://dx.doi.org/10.1177/1356389010380001>. [118]
- Jacob, S., S. Speer and J. Furubo (2015), “The institutionalization of evaluation matters: Updating the International Atlas of Evaluation 10 years later”, *Evaluation*, Vol. 21/1, pp. 6-31, <http://dx.doi.org/10.1177/1356389014564248>. [63]
- Johnson, K. et al. (2009), “Research on Evaluation Use A Review of the Empirical Literature From 1986 to 2005”, <http://dx.doi.org/10.1177/1098214009341660>. [167]
- King, J. et al. (2001), *Toward a Taxonomy of Essential Evaluator Competencies*. [149]
- Kothari, A. et al. (2009), “Is research working for you? validating a tool to examine the capacity of health organizations to use research”, *Implementation Science*, Vol. 4/1, p. 46, <http://dx.doi.org/10.1186/1748-5908-4-46>. [181]
- Kroll, A. and D. Moynihan (2018), “The Design and Practice of Integrating Evidence: Connecting Performance Management with Program Evaluation”, *Public Administration Review*, Vol. 78/2, pp. 183-194, <http://dx.doi.org/10.1111/puar.12865>. [17]
- Kusters, C. (2011), *Making evaluations matter: a practical guide for evaluators MSP book View project Strengthening Managing for Impact (SMIP) View project*, <https://www.researchgate.net/publication/254840956>. [103]

- Kusters, C. et al. (2011), "Making evaluations matter: a practical guide for evaluators", Centre for Development Innovation, Wageningen University & Research centre., [125]
<https://www.researchgate.net/publication/254840956>.
- Langer, L., J. Tripney and D. Gough (2016), *The science of using science: researching the use of Research evidence in decision-making*. [175]
- Lazaro, B. (2015), *Comparative Study on the Institutionalization of Evaluation in Europe and Latin America*, Eurosocal Programme. [5]
- Lázaro, B. (2015), *Comparative study on the institutionalisation of evaluation in Europe and Latin America*, Eurosocal Programme, Madrid, http://sia.eurosocal-ii.eu/files/docs/1456851768-E_15_ENfin.pdf (accessed on 9 July 2019). [54]
- Ledermann, S. (2012), "Exploring the Necessary Conditions for Evaluation Use in Program Change", *American Journal of Evaluation*, Vol. 33/2, pp. 159-178, [57]
<http://dx.doi.org/10.1177/1098214011411573>.
- Ledermann, S. (2012), "Exploring the Necessary Conditions for Evaluation Use in Program Change", <http://dx.doi.org/10.1177/1098214011411573>. [163]
- Leviton, L. and E. Hughes (1981), *Research on the Utilization of Evaluations: A Review and Synthesis view and Synthesis*. [100]
- Little, B. (ed.) (1979), *Speaking Truth to Power: The Art and Craft of Policy Analysis*. [135]
- Liverani, M., B. Hawkins and J. Parkhurst (2013), *Political and institutional influences on the use of evidence in public health policy. A systematic review*., [68]
<http://dx.doi.org/10.1371/journal.pone.0077404>.
- Mackay, K. (2007), *How to Build M&E Systems to Support Better Government*, The World Bank, [58]
<http://dx.doi.org/10.1596/978-0-8213-7191-6>.
- Maeda, A., M. Harrit and S. Mabuchi (2012), *Human Development Creating Evidence for Better Health Financing Decisions A Strategic Guide for the Institutionalization of National Health Accounts*, The World Bank, Washington, DC, <http://dx.doi.org/10.1596/978-0-8213-9469-4>. [69]
- Malčík, M. and A. Seberová (2010), "Meta-evaluation and Quality Standard of Final Evaluation Report. The New Educational Review", *The New Educational Review*, Vol. 22, pp. 149-164. [140]
- McDavid, J., I. Huse and L. Hawthorn (2006), *Program Evaluation and Performance Measurement: An Introduction to Practice*, <https://study.sagepub.com/mcdavid3e> (accessed on 28 January 2020). [10]
- Mcguire, M. and R. Zorzi (2005), "EVALUATOR COMPETENCIES AND PERFORMANCE DEVELOPMENT", *The Canadian Journal of Program Evaluation*, Vol. 20/2, pp. 73-99. [146]
- Mergaert, L. and R. Minto (2015), "Ex Ante and Ex Post Evaluations: Two Sides of the Same Coin? The Case of Gender Mainstreaming in EU Research Policy", *Symposium on Policy Evaluation in the EU*, <http://dx.doi.org/10.1017/S1867299X0000427X>. [37]
- Mideplan (2018), *Guide for Terms of References (Guía de Términos de Referencia)*, <https://documentos.mideplan.go.cr/share/s/DVyxte00R3a0T6E2QbfQww> (accessed on 21 August 2019). [126]

- Mideplan (2018), *Guide for the use of evaluations: guidelines for its implementation and follow-up on recommendations*,
<https://documentos.mideplan.go.cr/share/s/DDVZ114kTjCsTAXihi5Kw> (accessed on 3 September 2019). [180]
- Mideplan (2018), *National Evaluation Policy*,
<https://documentos.mideplan.go.cr/share/s/Ymx1WmMJTOWe9YyjieCHKQ> (accessed on 20 September 2019). [83]
- Ministerio de Desarrollo Social y Familia de Chile (2019), *Evaluación Social Ex Ante*,
<http://sni.ministeriodesarrollosocial.gob.cl/evaluacion-iniciativas-de-inversion/evaluacion-ex-ante/> (accessed on 5 July 2019). [35]
- Ministerio de Planificación Nacional y Política Económica (2018), *Manual de Evaluación para Intervenciones Públicas*,
<https://documentos.mideplan.go.cr/share/s/6eepeLCESrKkft6Mf5SToA> (accessed on 5 August 2019). [24]
- Ministry of Finance (2006), *The Norwegian Government Agency for Financial Management (DFØ)*, <https://www.regjeringen.no/en/dep/fin/about-the-ministry/etater-og-virksomheter-under-finansdepartementet/subordinateagencies/the-norwegian-government-agency-for-fina/id270409/> (accessed on 12 July 2019). [38]
- Ministry of Finance (Lithuania) (2011), *Recommendations on Implementation of Programs Evaluation Methodology*,
https://finmin.lrv.lt/uploads/finmin/documents/files/LT_ver/Veiklos_sritys/Veiklos_efektyvumo_tobulinimas/PVrekomendacijos2011.pdf (accessed on 27 August 2019). [132]
- Ministry of Finance of The Netherlands (2018), *Arrangements for periodic evaluation research*,
<https://wetten.overheid.nl/BWBR0040754/2018-03-27> (accessed on 12 July 2019). [6]
- Ministry of Internal Affairs and Communications (2017), “Basic Guidelines for Implementing Policy Evaluation (Revised)”, https://www.soumu.go.jp/main_content/000556221.pdf (accessed on 16 September 2019). [76]
- Ministry of Internal and Communications (2017), *Basic Guidelines for Implementing Policy Evaluation Revised*, http://www.soumu.go.jp/main_content/000556221.pdf. [190]
- Morton, M. (2009), *Applicability of Impact Evaluation to Cohesion Policy 1 Report Working Paper of*, https://ec.europa.eu/regional_policy/archive/policy/future/pdf/4_morton_final-formatted.pdf (accessed on 8 August 2019). [49]
- National Audit Office (GBR) (2013), *Evaluation in Government*,
<https://www.nao.org.uk/report/evaluation-government/> (accessed on 22 August 2019). [154]
- Neuhoff, A. et al. (2015), *The What Works Marketplace Helping Leaders Use Evidence to Make Smarter Choices 2 Invest in What Works Policy Series*, The Bridgespan group,
<http://www.results4america.org>. [177]
- Newman, J., A. Cherney and B. Head (2017), “Policy capacity and evidence-based policy in the public service”, *Public Management Review*, Vol. 19/2, pp. 157-174,
<http://dx.doi.org/10.1080/14719037.2016.1148191>. [164]

- Newman, K., C. Fisher and L. Shaxson (2012), "Stimulating Demand for Research Evidence: What Role for Capacity-building?", *IDS Bulletin*, Vol. 43/5, pp. 17-24, <http://dx.doi.org/10.1111/j.1759-5436.2012.00358.x>. [111]
- Newman, K., C. Fisher and L. Shaxson (2012), "Stimulating Demand for Research Evidence: What Role for Capacity-building?", *IDS Bulletin*, Vol. 43/5, pp. 17-24, <http://dx.doi.org/10.1111/j.1759-5436.2012.00358.x>. [183]
- OECD (2020), *Building capacity for evidence informed policy making*, OECD, Paris. [99]
- OECD (2020), *The Irish Government Economic and Evaluation Service: Using Evidence-Informed Policy Making to Improve Performance*, OECD Public Governance Reviews, OECD Publishing, Paris, <https://dx.doi.org/10.1787/cdda3cb0-en>. [98]
- OECD (2019), *Budgeting and Public Expenditures in OECD Countries 2019*, OECD Publishing, Paris, <https://dx.doi.org/10.1787/9789264307957-en>. [12]
- OECD (2019), *Evaluating Public Sector Innovation Support or hindrance to innovation?*, Observatory of Public Sector Innovation-OPSI, Paris, <https://oecd-opsi.org/wp-content/uploads/2019/05/Evaluating-Public-Sector-Innovation-Part-5a-of-Lifecycle-Report.pdf> (accessed on 11 September 2019). [59]
- OECD (2019), "Open Government Data Report". [107]
- OECD (2019), *Open Government in Biscay*, OECD Public Governance Reviews, OECD Publishing, Paris, <https://dx.doi.org/10.1787/e4e1a40c-en>. [11]
- OECD (2018), *Building Capacity for Evidence Informed Policy Making: Towards a Baseline Skill Set*, <http://www.oecd.org/gov/building-capacity-for-evidence-informed-policy-making.pdf> (accessed on 3 September 2019). [1]
- OECD (2018), *Centre Stage 2- The organisation and functions of the centre of government in OECD countries*, <https://www.oecd.org/gov/centre-stage-2.pdf>. [93]
- OECD (2018), *Cost-Benefit Analysis and the Environment: Further Developments and Policy Use*, OECD Publishing, Paris, <https://dx.doi.org/10.1787/9789264085169-en>. [47]
- OECD (2018), *Draft Policy Framework on Sound Public Governance*, <http://www.oecd.org/gov/draft-policy-framework-on-sound-public-governance.pdf> (accessed on 8 July 2019). [2]
- OECD (2018), "OECD Best Practice Principles for Regulatory Policy: Reviewing the Stock of Regulation". [123]
- OECD (2018), *OECD Performance Budgeting Survey*. [192]
- OECD (2018), *OECD Regulatory Policy Outlook 2018*, OECD Publishing, Paris, <https://dx.doi.org/10.1787/9789264303072-en>. [172]
- OECD (2017), *Governing Better Through Evidence-Informed Policy Making Options for an OECD Work Agenda*. [112]
- OECD (2017), *Government at a Glance - OECD*, OECD, Paris, <http://www.oecd.org/gov/govataglance.htm> (accessed on 9 July 2019). [4]

- OECD (2017), *Making policy evaluation work*. [105]
- OECD (2017), *Policy Advisory Systems: Supporting Good Governance and Sound Public Decision Making*, OECD Public Governance Reviews, OECD Publishing, Paris, <https://dx.doi.org/10.1787/9789264283664-en>. [186]
- OECD (2017), *Systems Approaches to Public Sector Challenges: Working with Change*, OECD Publishing, Paris, <https://dx.doi.org/10.1787/9789264279865-en>. [53]
- OECD (2016), *Engaging Public Employees for a High-Performing Civil Service*, OECD Public Governance Reviews, OECD Publishing, Paris, <https://dx.doi.org/10.1787/9789264267190-en>. [184]
- OECD (2016), *Evaluation Systems in Development Co-operation: 2016 Review*, OECD Publishing, Paris, <https://dx.doi.org/10.1787/9789264262065-en>. [144]
- OECD (2016), *Open Government: The Global Context and the Way Forward*, OECD Publishing, Paris, <https://dx.doi.org/10.1787/9789264268104-en>. [43]
- OECD (2016), *Supreme Audit Institutions and Good Governance: Oversight, Insight and Foresight*, OECD Public Governance Reviews, OECD Publishing, Paris, <https://doi.org/10.1787/9789264263871-en>. [21]
- OECD (2014), “Budget Review: Germany”, *OECD Journal on Budgeting*, Vol. 2. [81]
- OECD (2014), “Centre Stage Driving Better Policies from the Centre of Government”, [http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=gov/pgc/mpm\(2014\)3&doclanguage=en](http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=gov/pgc/mpm(2014)3&doclanguage=en) (accessed on 23 September 2019). [92]
- OECD (2014), *Chile’s Supreme Audit Institution: Enhancing Strategic Agility and Public Trust*, OECD Public Governance Reviews, OECD Publishing, Paris, <https://dx.doi.org/10.1787/9789264207561-en>. [96]
- OECD (2011), “Government at a Glance”. [174]
- OECD (2011), “Typology and implementation of spending reviews”, *OECD SBO Meeting on Performance and Results*, [http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=GOV/PGC/SBO\(2011\)9&doclanguage=en](http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=GOV/PGC/SBO(2011)9&doclanguage=en) (accessed on 1 August 2019). [13]
- OECD (2010), “DAC Guidelines and Reference Series: Quality Standards for Development Evaluation”, <https://www.oecd.org/development/evaluation/qualitystandards.pdf> (accessed on 9 July 2019). [48]
- OECD (2010), *Good Practices in Supporting Supreme Audit Institutions*, <http://www.oecd.org/dac/effectiveness/Final%20SAI%20Good%20Practice%20Note.pdf>. [95]
- OECD (2008), *Performance Budgeting: A Users’ Guide*, <https://www.oecd.org/gov/budgeting/Performance-Budgeting-Guide.pdf> (accessed on 2 August 2019). [19]
- OECD (forthcoming), *Ensuring the Good Governance of Evidence, taking stock of standards for policy design, implementation and evaluation*, OECD, Paris. [106]

- OECD-DAC (2009), "Guidelines for Project and Programme Evaluations", [84]
https://www.entwicklung.at/fileadmin/user_upload/Dokumente/Projektentwicklung/Englisch/Guidelines_for_Project_and_Programme_Evaluations.PDF (accessed on 20 September 2019).
- OECD-DAC (2002), *Glossary of Key Terms in Evaluation and Results Based Management*, [26]
<https://www.oecd.org/dac/evaluation/18074294.pdf> (accessed on 18 June 2019).
- Office fédéral de la justice (2005), *Guide for the evaluation of the efficacy of the Confederation (Guide de l'évaluation de l'efficacité à la Confédération)*, [137]
http://www.ofj.admin.ch/ejpd/fr/home/themen/staat_und_buerger/ref_evaluation/ref_umsetzung_art.html (accessed on 27 August 2019).
- Office of Management and Budget (2020), *MB-20-12 Program Evaluation Standards and Practices*, [86]
<https://www.whitehouse.gov/wp-content/uploads/2020/03/M-20-12.pdf>.
- Office of Management and Budget (2018), *Monitoring and Evaluation Guidelines for Federal Departments and Agencies that Administer United States Foreign Assistance*, [136]
<https://www.whitehouse.gov/wp-content/uploads/2017/11/M-18-04-Final.pdf> (accessed on 27 August 2019).
- Olejniczak, K. and E. Raimondo (2016), "Evaluation units as knowledge brokers: Testing and calibrating an innovative framework", *Evaluation*, Vol. 22/2, pp. 168-189, [182]
<http://dx.doi.org/10.1177/1356389016638752>.
- Olejniczak, K., E. Raimondo and T. Kupiec (2016), "Evaluation units as knowledge brokers: Testing and calibrating an innovative framework", *Evaluation*, Vol. 22/2, pp. 168-189, [64]
<http://dx.doi.org/10.1177/1356389016638752>.
- Oliver, K. et al. (2015), "Identifying public health policymakers' sources of information: comparing survey and network analyses", *The European Journal of Public Health*, Vol. 27/suppl_2, p. ckv083, [169]
<http://dx.doi.org/10.1093/eurpub/ckv083>.
- OMB (2010), *Section 200 - Overview of the Federal Performance Framework*. [178]
- Operational and Evaluation Audit Division of Costa Rica (2014), *INFORME DE AUDITORÍA DE CARÁCTER ESPECIAL SOBRE LOS PROCESOS DE SEGUIMIENTO, EVALUACIÓN Y RENDICIÓN DE CUENTAS PÚBLICA EN*, [156]
https://cgrfiles.cgr.go.cr/publico/jaguar/sad_docs/2015/DFOE-SAF-IF-09-2014-Recurrido.pdf (accessed on 22 August 2019).
- Parkhurst, J. (2017), *The politics of evidence : from evidence-based policy to the good governance of evidence*, Routledge, London, <http://researchonline.lshtm.ac.uk/3298900/> (accessed on 23 November 2018). [32]
- Patton, M. (1978), "Utilization-focused evaluation". [102]
- Picciotto, R. (2013), "Evaluation Independence in Organizations", *Journal of MultiDisciplinary Evaluation*, Vol. 9/20, p. 15. [56]
- Picciotto, R. (n.d.), *The Value of Evaluation Standards: A Comparative Assessment*, [120]
<http://evaluation.wmich.edu/jmde/Articles>.

- Pleger, L. and S. Hadorn (2018), "The big bad wolf's view: The evaluation clients' perspectives on independence of evaluations", *Evaluation*, Vol. 24/4, pp. 456-474, <http://dx.doi.org/10.1177/1356389018796004>. [115]
- Pleger, L. and S. Hadorn (2018), "The big bad wolf's view: The evaluation clients' perspectives on independence of evaluations", *Evaluation*, Vol. 24/4, pp. 456-474, <http://dx.doi.org/10.1177/1356389018796004>. [138]
- Podems, D. (2013), "Evaluator competencies and professionalizing the field: Where are we now?", *Canadian Journal of Program Evaluation*, Vol. 28/3, pp. 127-136. [150]
- Poder Ejecutivo Nacional (2018), *Decreto 292/2018: Evaluación de Políticas y Programas Sociales*, 11-04-2018, <http://servicios.infoleg.gob.ar/infolegInternet/verNorma.do?id=308653> (accessed on 8 July 2019). [33]
- Poder Ejecutivo Nacional de Argentina (2018), *Decreto 292/2018: Evaluación de Políticas y Programas Sociales*. [22]
- Polish Ministry of Infrastructure and Development (2015), *Guidelines of cohesion policy evaluation for period 2014-2020*, http://www.ewaluacja.gov.pl/media/13209/wytyczne_090915_final.pdf (accessed on 22 August 2019). [145]
- Punton, M. et al. (2016), *HOW CAN CAPACITY DEVELOPMENT PROMOTE EVIDENCE-INFORMED POLICY MAKING? Literature Review for the Building Capacity to Use Research Evidence (BCURE) Programme*, <http://www.itad.com/knowledge-and-resources/bcure> (accessed on 6 September 2019). [185]
- Results for America (2017), *Government Mechanisms to Advance the Use of Data and Evidence in Policymaking: A Landscape Review*. [166]
- Results for America (2017), *Government Mechanisms to Advance the Use of Data and Evidence in Policymaking: A Landscape Review*. [193]
- Robert Picciotto (2013), "Evaluation Independence in Organizations", *Journal of MultiDisciplinary Evaluation*, Vol. 9/20, p. 15. [114]
- Robinson, M. (2014), *Spending reviews*, <http://www.pfmresults.com>. (accessed on 25 June 2019). [187]
- Roh, J. (2018), "Improving the government performance management system in South Korea", *Asian Education and Development Studies*, Vol. 7/3, pp. 266-278, <http://dx.doi.org/10.1108/AEDS-11-2017-0112>. [79]
- Rutter, J. (2012), *Evidence and Evaluation in Policy making*, Institute for Government, https://www.instituteforgovernment.org.uk/sites/default/files/publications/evidence%20and%20evaluation%20in%20template_final_0.pdf. [52]
- Schillemans, T. and M. Bovens (2011), *The Challenge of Multiple Accountability: Does Redundancy lead to Overload?*. [62]
- Schoenefeld, J. and A. Jordan (2017), "Governing policy evaluation? Towards a new typology", *Evaluation*, Vol. 23/3, pp. 274-293, <http://dx.doi.org/10.1177/1356389017715366>. [39]

- Scriven, M. (1969), "An introduction to meta-evaluation", *Educational Product Report*, Vol. 2. [139]
- Secretaria de Desarrollo Social (2015), *Decree for which the Council of social Development Policy Evalaution is regulated..* [91]
- Shaxson, L. (2019), "Uncovering the practices of evidence-informed policy-making", *Public Money & Management*, Vol. 39/1, pp. 46-55, <http://dx.doi.org/10.1080/09540962.2019.1537705>. [108]
- Sinatra, G., D. Kienhues and B. Hofer (2014), "Educational Psychologist Addressing Challenges to Public Understanding of Science: Epistemic Cognition, Motivated Reasoning, and Conceptual Change", <http://dx.doi.org/10.1080/00461520.2014.916216>. [113]
- Smismans, S. (2015), "Policy Evaluation in the EU: The Challenges of Linking Ex Ante and Ex Post Appraisal", *Symposium on Policy Evaluation in the EU*, <http://dx.doi.org/10.1017/S1867299X00004244>. [14]
- Smismans, S. (2015), "Policy Evaluation in the EU: The Challenges of Linking Ex Ante and Ex Post Appraisal", *Symposium on Policy Evaluation in the EU*, <http://dx.doi.org/10.1017/S1867299X00004244>. [34]
- Stern, E., M. Saunders and N. Stame (2015), "Standing back and looking forward: Editors' reflections on the 20th Anniversary of Evaluation", *Evaluation*, Vol. 21/4, pp. 380-390, <http://dx.doi.org/10.1177/1356389015608757>. [101]
- Steuerle, E. and L. Jackson (2016), *Advancing the power of economic evidence to inform investments in children, youth, and families*, National Academies Press, <http://dx.doi.org/10.17226/23481>. [46]
- Stevahn, L. et al. (2005), "Establishing Essential Competencies for Program Evaluators", *ARTICLE American Journal of Evaluation*, <http://dx.doi.org/10.1177/1098214004273180>. [147]
- Stufflebeam, D. (2001), *Method Notes Evaluation Checklists: Practical Tools for Guiding and Judging Evaluations*, <http://www.wmich.edu/evalctr/checklists/>. [143]
- Stufflebeam, D. (1978), "Meta evaluation: an overview", *Evaluation and The Health Professions*, Vol. 1/1, <https://journals.sagepub.com/doi/pdf/10.1177/016327877800100102> (accessed on 19 August 2019). [141]
- Superu (2018), *Making sense of evidence: A guide to using evidence in policy*, <https://thehub.sia.govt.nz/assets/Uploads/Making-Sense-of-Evidence-handbook-FINAL.pdf> (accessed on 3 September 2019). [179]
- Swiss Federal Audit Office (2019), *Involving Stakeholders in Evaluation at the Swiss Federal Audit Office*, http://www.program-evaluation.ccomptes.fr/images/stories/evenements/Vilnius_2019/Presentation_Switzerland_Stakeholder_Involvement_WGEPPP_2019.pdf (accessed on 6 September 2019). [189]
- The Cabinet Secretariat (2019), *The status-quo about the promotion of statistics reform*, <http://www.kantei.go.jp/jp/singi/toukeikaikaku/dai5/siryou1.pdf> (accessed on 2 September 2019). [74]

- The Committee on Promoting EBPM (2017), *Guidelines on securing and developing human resources for the promotion of EBPM*, <https://www.gyokaku.go.jp/ebpm/img/guideline1.pdf> (accessed on 2 September 2019). [73]
- The European Network for Rural Development (2014), *The Ex Ante Evaluation 2014-2020 RDPs*, https://enrd.ec.europa.eu/evaluation/publications/guidelines-ex-ante-evaluation-2014-2020-rdps_en (accessed on 29 January 2020). [36]
- The Ministry of Internal Affairs and Communication (2005), *Policy Evaluation Implementation Guidelines*, https://www.soumu.go.jp/main_content/000556222.pdf. [78]
- The Ministry of Internal Affairs and Communications (2010), “Guidelines for Publication of Information on Policy Evaluation”, http://www.soumu.go.jp/main_content/000556224.pdf (accessed on 16 September 2019). [77]
- The Statistical Reform Promotion Council (2017), *The final report of the Statistical Reform Promotion Council. (In Japanese)*, http://www.kantei.go.jp/jp/singi/toukeikaikaku/pdf/saishu_honbun.pdf (accessed on 2 September 2019). [72]
- The World Bank (2018), *Spending Review Manual: Bulgaria*. [15]
- Treasury Board Secretariat (2019), *Evaluation in the Government of Canada - Canada.ca*, <https://www.canada.ca/en/treasury-board-secretariat/services/audit-evaluation/evaluation-government-canada.html> (accessed on 20 September 2019). [87]
- Treasury Board Secretariat (2013), *Assessing program resource utilization when evaluating federal programs*, <https://www.canada.ca/en/treasury-board-secretariat/services/audit-evaluation/centre-excellence-evaluation/assessing-program-resource-utilization-evaluating-federal-programs.html> (accessed on 2 August 2019). [88]
- Treasury Board Secretariat (2010), *Supporting Effective Evaluations: A Guide to Developing Performance Measurement Strategies*, <https://www.canada.ca/en/treasury-board-secretariat/services/audit-evaluation/centre-excellence-evaluation/guide-developing-performance-measurement-strategies.html> (accessed on 2 August 2019). [89]
- United Nations Evaluation Group (2016), *Norms and Standards for Evaluation*. [121]
- United States Office of Management and Budget (2019), *Phase 1 Implementation of the Foundations for Evidence-Based Policymaking Act of 2018: Learning Agendas, Personnel, and Planning Guidance*, <https://www.whitehouse.gov/wp-content/uploads/2019/07/M-19-23.pdf>. [75]
- Vaessen, J. (2018), *New blogpost - Five ways to think about quality in evaluation*, <https://www.linkedin.com/pulse/new-blogpost-five-ways-think-quality-evaluation-jos-vaessen> (accessed on 21 June 2019). [104]
- Vammalle, C. and A. Ruiz Rivadeneira (2017), “Budgeting in Chile”, *OECD Journal on Budgeting*, Vol. 16/3, <https://dx.doi.org/10.1787/budget-16-5jfw22b3c0r3>. [44]
- van Ooijen, C., B. Ubaldi and B. Welby (2019), “A data-driven public sector: Enabling the strategic use of data for productive, inclusive and trustworthy governance”, *OECD Working Papers on Public Governance*, No. 33, OECD Publishing, Paris, <https://dx.doi.org/10.1787/09ab162c-en>. [71]

- Viñuela, L., D. Ortega and F. Gomes (2015), *Technical Note - Mechanisms and incentives for the adoption of evaluation of Policies and Programs to improve the Efficiency of Public Expenditure*. [66]
- Walker, K. and K. Moore (2011), *Performance Management and Evaluation: What's the difference?*, <https://www.childtrends.org/wp-content/uploads/2013/06/2011-02PerformMgmt.pdf> (accessed on 18 July 2019). [16]
- Weiss, C. (1993), "Where politics and evaluation research meet", *Evaluation Practice*, Vol. 14/1, pp. 93-106, [http://dx.doi.org/10.1016/0886-1633\(93\)90046-R](http://dx.doi.org/10.1016/0886-1633(93)90046-R). [40]
- Weiss, C. and C. Weiaa Harvard (1998), "Have We Learned Anything New About the Use of Evaluation?", *American Journal of Evaluation*, Vol. 19/1, pp. 21-33. [160]
- World Bank et al. (2019), *World Bank Group Evaluation Principles*, <http://www.worldbank.org>. [124]
- Yang, S. and A. Torneo (2016), "Government Performance Management and Evaluation in South Korea: History and Current Practices", *Public Performance & Management Review*, Vol. 39/2, pp. 279-296, <http://dx.doi.org/10.1080/15309576.2015.1108767>. [80]
- Zida, A. et al. (2017), "Evaluating the Process and Extent of Institutionalization: A Case Study of a Rapid Response Unit for Health Policy in Burkina Faso.", *International journal of health policy and management*, Vol. 7/1, pp. 15-26, <http://dx.doi.org/10.15171/ijhpm.2017.39>. [70]

Annex A. Data and methodology

This comparative analysis of policy evaluation builds on the data collected through the 018 OECD Survey on Policy Evaluation (hereafter, the “OECD Survey”). The OECD Survey is a direct response to the request to collect better data in the area of evidence informed policy making by the Public Governance Committee, in the context of the development of a Policy Framework on Sound Public Governance. (see <https://www.oecd.org/governance/policy-framework-on-sound-public-governance>). This Framework was derived from the aim to support policymakers by consolidating an integrated vision and coherent narrative of the main elements of sound public governance. The information on policy evaluation practices was gathered as part of the effort to inform the development of the Framework chapters on policy formulation, implementation and evaluation.

While the report mainly draws on the data gathered through this OECD Survey on Policy Evaluation, it also uses related data from the OECD performance budgeting survey (<https://www.oecd.org/governance/budgeting/internationalbudgetpracticesandproceduresdatabase.htm>) and also from the work on measuring regulatory governance, under the Regulatory Policy Committee (<http://www.oecd.org/gov/regulatory-policy/measuring-regulatory-performance.htm>). It also relies on comparative data on the Centres of Government (COG) from the “Centre Stage” Publications and related COG surveys (see <https://www.oecd.org/gov/centre-stage-2.pdf>).

After a consultation and approval by the Public Governance Committee Delegates, and a piloting with a few volunteer countries, the survey was officially launched in mid 2018, with a range of countries involved in the OECD work on public governance. The process of data collection ended finally in mid 2019, with data being available for 42 countries. Countries were allowed to update the results to show any reforms in place and implemented as of mid 2019.

The survey is structured in two parts:

- Policy evaluation across government.
- Policy evaluation at the sector/thematic level.

In the first part of the survey, respondents were asked to provide information and data from a whole of government perspective on policy evaluation. The questionnaire was sent to the main institution responsible for the promotion, coordination and implementation of policy evaluation across government and answered by civil servants at director/assistant secretary level. The current report mainly draws from this main part of the survey.

The survey had a second component at sector and thematic level where respondents from the ministries of health and public sector reform were asked to provide information and data about the governance of policy evaluation in their institution. The current report has done a preliminary use of the data for the purpose of highlighting variations between main practices and some of the sectoral practices.

Survey Responses

Data were received from 42 countries (including 35 OECD member countries with the exception of Luxembourg). More specifically, 42 countries replied to the first part of the OECD Survey, and 33 countries’ ministries of health or equivalent institution and 31 ministries of public sector reform or equivalent institution replied to the second part of the survey. In terms of the non-members at the time of running the survey, responses were received from Argentina, Brazil, Bulgaria, Colombia, Costa Rica, Kazakhstan, and Romania.

The OECD Survey questionnaire is available at the link below: www.oecd.org/gov/institutionalisation-of-policy-evaluation-questionnaire.pdf.

Throughout the report, data are presented according to countries, which have been abbreviated according to the official ISO country codes.

Table A.1. ISO codes of the countries referred to in this report

1	Argentina	ARG	22	Israel	ISR
2	Australia	AUS	23	Italy	ITA
3	Austria	AUT	24	Japan	JPN
4	Belgium	BEL	25	Kazakhstan	KAZ
5	Bulgaria	BGR	26	Korea	KOR
6	Brazil	BRA	27	Latvia	LVA
7	Canada	CAN	28	Lithuania	LTU
8	Chile	CHL	29	Mexico	MEX
9	Colombia	COL	30	Netherlands	NDL
10	Costa Rica	CRI	31	New Zealand	NZL
11	Czech Republic	CZE	32	Norway	NOR
12	Denmark	DNK	33	Poland	POL
13	Estonia	EST	34	Portugal	PRT
14	Finland	FIN	35	Romania	ROU
15	France	FRA	36	Slovakia	SVK
16	Germany	DEU	37	Slovenia	SVN
17	Great Britain	GBR	38	Spain	ESP
18	Greece	GRC	39	Sweden	SWE
19	Hungary	HUN	40	Switzerland	CHE
20	Iceland	ISL	41	Turkey	TUR
21	Ireland	IRL	42	United States	USA

Process for data collection and validation

The survey was distributed through the network of the Public Governance Committee Delegates, who were invited to designate a focal point at national level. The Secretariat carefully reviewed all the survey submissions, with additional requests for information on approaches, clarifications and potential good practices. A process of validation and mutual dialogue with each of the responding country did follow. Moreover, an experts' meeting was organised in September 2019 where the survey respondents were invited to the OECD to discuss preliminary results and given a final opportunity to validate their country's data. (See <http://www.oecd.org/gov/agenda%20expert%20group%20meeting%20final.pdf>).

The survey questions were designed to be responded to in a quantifiable manner, such as yes or no options and multiple-choice questions. Whenever possible, the OECD Survey offered space for countries to give additional information. While the survey can provide aggregate comparative information, additional boxes and qualitative analysis were added to reflex the complex processes of policy evaluation and evidence informed policy making.

OECD Public Governance Reviews

Improving Governance with Policy Evaluation

LESSONS FROM COUNTRY EXPERIENCES

Policy evaluation is a critical element of good governance, as it promotes public accountability and contributes to citizens' trust in government. Evaluation helps ensure that decisions are rooted in trustworthy evidence and deliver desired outcomes. Drawing on the first significant cross-country survey of policy evaluation practices covering 42 countries, this report offers a systemic analysis of the institutionalisation, quality and use of evaluation across countries and looks at how these three dimensions interrelate. The report also covers cross-cutting aspects related to regulatory assessment and performance budgeting. The analysis illustrates the role and functions of key institutions within the executive, such as centres of government and ministries of finance. It also underlines the role of supreme audit institutions.



PRINT ISBN 978-92-64-68184-2

PDF ISBN 978-92-64-82982-4



9 789264 681842