Policy & Politics • vol 48 • no 1 • 89–110 • © Policy Press 2020
Print ISSN 0305-5736 • Online ISSN 1470-8442 • https://doi.org/10.1332/030557319X15579230420108
Accepted for publication 24 April 2019 • First published online 04 September 2019

SPECIAL ISSUE • Policy-making as designing: the added value of design thinking for public administration and public policy

article

Policy labs: the next frontier of policy design and evaluation?

Karol Olejniczak, karol.olejniczak@swps.edu.pl
SWPS University of Social Sciences and Humanities, Poland

Sylwia Borkowska-Waszak, sylwia.borkowska@strath.ac.uk
University of Strathclyde, UK

Anna Domaradzka-Widła, anna.domaradzka@gmail.com
University of Warsaw, Poland

Yaerin Park, yrpark7@gwu.edu
George Washington University, USA

The article explores the potential benefits to public policy of combining traditional evaluative inquiry with insights developed dynamically in policy labs. Twenty leading labs from five continents are critically analysed through a literature review as well as policy and programme evaluation practices, assessing the extent to which the purpose, structures and processes used in policy labs address three challenges: (1) establishing the causality and value of public interventions, (2) explaining mechanisms of change, and (3) utilising research findings in public policy. The article concludes that creating synergies between evaluation inquiry and policy labs can improve the design and implementation of public policy and programmes.

key words policy labs • policy innovation • evaluation practice • policy design • public policy

To cite this article: Olejniczak, K., Borkowska-Waszak, S., Domaradzka-Widła, A. and Park, Y. (2020) Policy labs: the next frontier of policy design and evaluation?, vol 48, no 1, 89–110, *Policy & Politics*, DOI: 10.1332/030557319X15579230420108

Introduction

Policy labs have been emerging all over the world with a mission to support policy practitioners with innovative solutions, grounded in empirical research. This trend started less than a decade ago (Price, 2015; Tõnurist et al, 2017), and a recent study has identified 78 policy labs in the European Union alone (Fuller and Lochard, 2016). While the idea of policy labs has gained considerable popularity, the term tends to be applied to different activities and approaches. This could be explained by the diverse strands that have provided methods for labs, such as design thinking, behavioural insights, collaborative governance and social entrepreneurship. This lack of a coherent typology of labs limits the possibility of comparing their operations and outcomes for academic and practical purposes.

The emerging practices of policy labs claim to bring promising innovations to the public sector. However, it remains unclear to what extent these varied practices provide truly new ways of addressing the challenges that have already been identified by existing public policy routines.

This article addresses the problem by looking at policy labs from the perspective of a well-established evaluation practice (Scriven, 1977; Chelimsky, 1985; Shaw et al, 2006; Newcomer et al, 2015). Evaluation has been systematically developing since the 1960s, accompanying subsequent waves of public programmes and becoming a standard of modern public management around the world (European Commission, 1995; Furubo et al, 2002; OECD, 1998; Jacob et al, 2015; Bohni et al, 2018). It has a dynamic worldwide community covering all policy fields, with numerous international, national and even regional associations across the world (for example, American Evaluation Association, European Evaluation Society, Australasian Evaluation Society) and its own highly respectable journals (for example, American Journal of Evaluation, Evaluation, Evaluation and Programme Planning). In order to underline the importance of evaluation practice for evidence-informed public policies and social betterment, the year 2015 was celebrated as an International Year of Evaluation (United Nations Evaluation Group, 2013).

Over the years, three broad challenges have emerged from global evaluation practice in relation to its contribution to public policies:

- *Establishing what solutions work* highlights the need of verifying the causal relation between public interventions and observed changes, and making a fair judgement about their value. This provides public policy with accountability for executed programmes, and indicates which interventions are worth financing in the future.
- *Explaining why solutions work (or not)* covers approaches and methods that provide insights into the complex nature of mechanisms responsible for socio-economic change. An understanding of these mechanisms enables the designing of more effective interventions.
- *Transferring research findings into policy actions* focuses on bridging the gap between the extensive production of policy studies and their limited utilisation in decision-making processes. Resolving this issue could substantially strengthen evidence-informed policies.

This article challenges the emerging practice of policy labs, by addressing the following question:

> What possible synergies between evaluation practice and policy labs could help in addressing the three challenges of public policy: establishing what interventions work, explaining their change mechanisms, and using research findings?

For this purpose, we analyse the functions, structures and processes of 20 well-established policy labs from Western Europe, North America, South America, Australia and Asia. The comparative, deductive content analysis covers grey literature, documents and web content on lab activities, as well as records from the analysed sample of policy labs.

The article consists of five parts. The next section provides a literature review explaining in more detail the nature of each of the aforementioned challenges of public policy that have emerged from evaluation practice. Section three provides the method of case study selection and the analytical framework used for reviewing and synthesising policy lab operations. Section four presents findings on the labs' functions, structures and their processes. These findings are confronted with the three challenges identified by evaluation practice, leading to a discussion of the possible advantages gained from synergies between evaluation inquiry and policy labs. The final section summarises the conclusion and indicates areas for further research.

The article advances our understanding of the synergies between policy labs and evaluation practice that could enhance evidence-informed policies. We find that policy labs are not an alternative development to established practices of public policy. They are rather a promising addition. The structures, functions and processes of policy labs can help to address evaluation challenges mainly related to explaining why solutions work (focusing on users, unpacking behavioural change mechanisms, building and testing solutions in iterative REACT process) and translating research findings into decision-making (involving policy stakeholders and providing platforms for safe learning through experimentation). At the same time, policy labs may benefit from evaluation theory and practice in gaining a deeper understanding of challenges that relate to establishing what works, as well as valuing public interventions, incorporating systems thinking and complexity, and scaling-up solutions. We believe that these synergies, when turned into an exchange of practices, could eventually enhance the effectiveness of public policies.

Challenges emerging from evaluation practice

Evaluation practice is an integral element of the policy process, which can be characterised as a series of stages of problem solving (from framing the policy issue, through development of solutions, selection of tools, implementation, to final assessment of effects) (Weiss, 1999; Howlett, 2011; Colebatch and Hoppe, 2018). Evaluation on the one hand ensures the accountability of public investments, and on the other, it enables learning and therefore the improved utility of public interventions over time (Chelimsky, 1997; Mark et al, 2000; Newcomer et al, 2015).

Single evaluation inquiries can be oriented on different functions, such as improving interventions planning, increasing the technical quality of performance, improving accountability, strengthening internal organisational capacities, enhancing ownership

among stakeholders, or even empowering marginalised groups (Batterbury, 2006; Mertens, 2008). However, the ultimate goal of evaluation is 'social betterment' through policy learning (Henry et al, 2003), since evidence-informed interventions are likely to be more effective in serving citizens.

At the level of everyday management practices, evaluation is defined as the use of a diverse spectrum of socio-economic research methods for systematic inquiry of the worth and merit of public interventions. It includes data collation, analysis, interpretation, assessment and dissemination of results (Patton, 2004; Mathison, 2005). Evaluation inquiries are executed by independent teams of social science researchers who follow ethical and methodological standards. The evaluation tasks are often assigned to designated teams of public professionals called 'evaluation units' who either execute studies by themselves, or commission studies and convey their findings to the institutions responsible for planning and implementing policies.

Evaluation practice has systematically developed since the 1960s, starting in the United States in response to the Great Society initiative. The large-scale federal programmes were coupled with a mandate for evaluation, often conducted in the form of pilot programmes and experiments (Patton, 2004) – similarly to policy labs today. Evaluation activities grew in popularity in the 1970s as part of the toolbox for rationalising public expenditures in the era of crisis. The performance-based management promoted by the OECD, World Bank and European Union made monitoring and evaluation a standard of modern public management. This further disseminated evaluation practice across the world (Derlien, 1990), although the policy scope and institutional embedding of evaluation vary across countries (Furubo et al, 2002; Jacob et al, 2015).

As the theory and practice of evaluative inquiry have advanced over the years, a number of challenges and major issues for debate have emerged. Alkin (2013) summarises these developments in the form of an evaluation tree. He identifies the shared roots of accountability, control and social inquiry, and then points out three branches: challenges in the philosophy of valuing, methodological issues, and the use of evaluation (Alkin, 2013). So far, this evaluation tree constitutes the most comprehensive attempt to map evaluation challenges; however, it focuses mainly on theoretical developments, and is largely US-centric.

A more practice-oriented approach has been encapsulated by the so-called evaluation realists, who articulate evaluation challenges in the form of a mission statement. Evaluation aims at establishing 'What works, for whom and in what context' (Pawson and Tilley, 1997). The first part of the question refers to the outcomes and measures of success. The other two search for mechanisms triggered by interventions in a target population and in a certain environment. In time, this question was linked to a broader evidence-based policy movement (Pawson, 2006; Davies et al, 2009; Shillabeer et al, 2011).

Following these debates, we have organised the evaluation challenges discussed in this article around three issues: establishing what policy solutions work, explaining why solutions work (or not), and transferring those research findings into policy actions. We discuss them in more detail in the next subsections.

The challenge of establishing what solution works

Value assessment of policy intervention through appropriate evaluation methods and practices has been at the centre of interest of both theorists and practitioners of public policy (Alkin, 2013; Dunn, 2017; Wildawsky, 2017). It is important for public

policy because it provides accountability for executed programmes, and indicates which interventions are worth financing in the future.

There are two main limitations of assessing the value of policy interventions. The first is the causality issue – the relation between an implemented solution and the observable change. Causality can be approached from various stances, such as the successionist, configurationist and generativist perspectives (Ray Pawson's classification discussed in Stern et al, 2012). While the successionists focus on analysing the direct association between a treatment and an observed outcome, the configurationists pay attention to the attributes that lead to variations in outcomes, and the generativists focus on identifying the supporting factors underlying an observed causal relation (Stern et al, 2012). In evaluation practices, researchers are expected to adopt various analytical approaches and designs depending on the causality stances and modes they intend to advocate.

The second challenge is related to the politics of valuing interventions, namely establishing evaluation criteria and reference points. The key criteria used in the assessment of public policy programmes and interventions include utility, legality, transparency, effectiveness, efficiency and sustainability (Olejniczak and Mazur, 2014). Depending on the criteria used, different pictures of policy outcomes emerge. Evaluators are aware of this issue, and they are cautious about clarifying what criteria can be applied. Additional issues in valuing interventions are created by the timeframe and inclusion of observable side effects. Evaluators analyse the short-term results, long-term impact and side effects of the policy intervention, and communicate the findings to the stakeholders. At times, the stakeholders involved in certain evaluation processes may require the outputs to be delivered within a shorter timeframe, while some actors prefer to see the results within a relatively longer timeframe. Thus, there are challenges and trade-offs associated with the politics of valuing interventions.

The challenge of explaining why a solution works or does not work

Public policy is about triggering a behavioural change mechanism (Shafir, 2013; Weaver, 2015; World Bank, 2015). This recent perspective is very much in line with the public policy 'classics' stating that public policy is about changing the choices and behaviours of policy actors (Lasswell, 1951; Simon, 1997; Wildawsky, 2017).

Evaluation literature follows this logic, arguing that public interventions (projects, programmes, policies and regulations) should be viewed as levers designed to activate certain change mechanisms among policy addressees, which in turn should lead to the desired effects (Astbury and Leeuw, 2010; Chen, 2004; Rossi et al, 1999).

This perspective, confronted with everyday policy practice, highlights two widespread shortcomings. First, designers of policy interventions often ignore the existence of these mechanisms, assuming a direct, automatic link between a policy action (input) and the policy addressees reaction (outcome), and neglecting side effects. This so-called 'black box' approach to policy design has been widely criticised in both practice and literature of evaluation (Astbury and Leeuw, 2010; Pawson, 2013).

Second, even when unpacking the black box of mechanisms, policy practitioners often follow the rational choice theory, assuming the full rationality of policy addressees, and their unchanging set of preferences (Amadae, 2007). This assumption stands in contrast with recent empirical findings of cognitive psychology, revealing how bounded rationality often leads to systematic errors and biases in decision-making (Simon, 1997; Kahneman, 2011; Shafir, 2013).

Therefore, in order to design effective policy it is crucial to obtain more realistic insights into the change mechanisms that drive the response of the policy addressees to the implemented policy measures (Shafir, 2013; Weaver, 2015).

The challenge of transferring research findings into policy actions

The *raison d'être* of all applied policy research, including evaluation studies, is to be used in decision-making processes. In its early days, evaluation practice was driven by an assumption, similar to the one made by the community of policy analysts, that the relation between the production of knowledge and its application is linear. Thus, the execution of credible studies alone should be a sufficient condition for its use in the policymaking process.

By the end of the 1970s, the policy research community started to recognise the growing gap between the production of analyses and their incorporation into programme decision-making (Weiss, 1988). By linking evaluation literature with the broader body of work on knowledge utilisation, three main causes can be identified. The first problem is the issue of timing. The majority of evaluations have a summative nature, thus they come at a relatively late stage in the policy cycle. But even ex-ante evaluations that aim to comment on programme drafts often arrive late in the decision-making process – when decisions have already been made and consensus around the programme has been built.

The second issue lies in the differences between the two communities – the producers of knowledge (researchers, analysts, evaluators), and the potential users of knowledge (decision-makers – politicians, senior civil servants, programme and project managers). The crucial discrepancies relate to different purposes, mindsets and languages. Policy researchers focus on valid knowledge generation and finding optimal solutions. Decision makers focus on finding a solution that allows the feasible amelioration of problems. Researchers tend to explore the details and increase uncertainty, while decision makers try to decrease the complexity and uncertainty they cope with on an everyday basis. Finally, evaluators tend to use academic language, focusing their narrative on theories and methodological specifics, while practitioners use the language of actions and practical implications (Caplan, 1979; Nutley et al, 2007).

The third issue is the complex nature of organisational learning combined with the peculiarities of public sector decision-making. Individual and organisational policy actors absorb information and learn in complex, non-linear ways (Leeuw et al, 1994; Olejniczak and Mazur, 2014). Furthermore, bureaucratic organisations have their own established routines, and tend to develop risk avoidance. Challenging the status quo of a particular policy requires a body of evidence aligned with other organisational and political processes that can eventually create a punctuated equilibrium. This makes the learning process highly incremental.

Scope and method of the study

There is still a scarcity of academic literature providing a worldwide overview on labs. Previous examples of research using case studies have applied a literature search, and selected labs mentioned in at least two different sources (McGann et al, 2018). As a result, European policy labs have dominated the selections, as they are more

present in English-speaking publications than Asian, African and South-American labs. Moreover, policy labs in Europe have operated for relatively longer, have had more opportunities to network, and are consequently more widely recognised.

In order to ensure higher generalisability of the findings, it is important to keep the balance between the selection of case studies from various geographical locations and the labs' level of operation (international, national, regional or local). Therefore, this study aimed at finding a worldwide database of policy labs, including their geographical location and operational level. The only sources providing such a broad overview available online, at the time of conducting the study, were the 'World of Labs' map prepared by NESTA (Price, 2015), updated by Irish academics (O'Rafferty, 2016), and the report *Public Policy Labs in European Union Member States* by the EU Policy Lab (Fuller and Lochard, 2016). Their non-academic and European origin could have been a drawback, possibly limiting their knowledge about the labs in other continents. However, NESTA provided information on at least five labs on each continent, while both organisations had prepared networking events for labs: NESTA's *Lab Works* in 2015 (London, UK) and 2016 (Santiago, Chile), and EU Policy Lab's *Lab Connections* in 2016 (Brussels, Belgium). Thus, they appeared up-to-date and reliable, and their consolidation formed the basis for further case study selection. The list was updated and validated using the review of data added to NESTA's webpage as well as desk research. Table 1 summarises the initial pool for case study selection.

The initial desk research revealed that the lab population is highly incoherent when it comes to experience, published online contents, and the relationship between their activities and public policies. Therefore, for the purpose of selecting a comparable pool of cases, further selection criteria were applied.

The first filter was the definition of the phenomenon. The study focused on policy labs – the entities that declare a public policy focus, for example, policy design, solutions for citizens or social impact, providing an open forum for new ideas and solutions to social problems. That excluded entrepreneurship labs, the majority of which support start-up companies, and 'influencers' – organisations that are interested in labs and social innovation, publish articles and reports about them, and offer networking events, but do not design public policies on their own. Examples of excluded entities are: NESTA (UK), iMinds (Belgium), OECD Observatory for Public Sector Innovation (France), World Bank (USA) and LabGov (Italy).

Table 1: Policy Labs worldwide

Policy Labs	*Local*	*Regional*	*National*	*Other*	*Total*
Europe	16	21	19	9	**65**
North America	15	17	8	4	**44**
Asia	1	2	6	6	**15**
Latin America	2	1	5	1	**9**
Africa	0	0	1	6	**7**
Australia & Oceania	1	0	3	2	**6**
TOTAL	**35**	**41**	**42**	**28**	**146**

Source: Elaborated by the authors on the basis of Fuller and Lochard, 2016; O'Rafferty, 2016; Price, 2015

The second filter was related to the geographical scope. The study aimed to include the continents with the highest number of nationally grown policy labs addressing a diversity of national, regional and local activities. This criterion led to the exclusion of Africa as it hosted mostly international (for example, UNESCO) labs by the time that this study was conducted (from May 2016 to May 2017). Moreover, Asia and Australia were merged into one geographical group. In order to avoid an imbalance (with Europe and North America dominating the population), we decided to have an equal number of case studies from all four geographical groups.

In order to identify the labs to be studied in each geographical area, the last sampling step was introduced – focusing on positive outliers. These were defined in terms of experience (labs in operation for at least one year and active at the time of the study), and data availability (labs having an operative website with information on their activities and additional documents, available in English or Spanish). We are aware of the limitations of a web-based study. However, we assumed that website communication is an important dissemination tool for policy labs. Therefore, in the case of positive outliers, the websites were sufficiently developed for us to gather reliable data, especially in terms of labs' know-how and experience. Whenever possible, we complemented webpage data with official documents produced by labs, such as reports, articles and guidebooks.

The final list of selected case studies covers 20 labs located on five continent groups, in 16 countries and 20 cities. It includes diverse labs in terms of all criteria on every continent. In each geographical group there are labs with various ownership statuses, levels of operation and years of experience (see Table 2).

Data on the selected labs was collected using desk research conducted by the authors. Analysis was based on exploratory coding followed by a comparative analysis. Codes were formed around three broad groups of issues:

1 Structures – who they are. This analysis covered organisational arrangements, legal status and form (public, private, NGOs), level of operation, experience in their activity as a lab, and their networks.
2 Functions – what labs do and what outcomes they produce. Analysis covered labs' overall mission, topics covered, types of interventions (regulations, small community projects, big programmes, and so on), and their reported impact, including case studies and the scope of influence on policies.
3 Processes – how they do it. This included the approaches, methods and tools that labs apply in designing public interventions.

The findings from the analysis were synthesised in the form of comparative tables, and these formed the basis for a discussion on labs' contribution to the main challenges identified by evaluation practice.

Findings: a review of lab operations and their potential contribution

This section presents the findings from the comparative study on policy labs and discusses how they could contribute to addressing the three policy challenges presented in the literature section. We summarise the key patterns that emerged across cases in relation to the structures, purpose and processes of policy labs. Due to a lack of

Table 2: Summary of selected case studies

	No.	*Policy lab*	*Location*	*Launch year*	*Legal status*	*Level of operation*
Europe	1	**Design Policy Lab**	Italy, Milan	2000	academic	local, international
	2	**Kennisland**	Netherlands, Amsterdam	1999	NGO	local
	3	**La 27e Région**	France, Paris	2008	NGO of public entities	regional
	4	**MindLab**	Denmark, Copenhagen	2002	public	national
	5	**UK Policy Lab**	UK, London	2014	public	national
North America	6	**Alberta CoLab**	Canada, Alberta	2014	public	regional, local
	7	**GovLab**	USA, New York	2012	academic	national, regional, local
	8	**GOVLabPHL**	USA, Philadelphia	2016	public–academic partnership	local
	9	**Ideas42 & Gov42**	USA, Chicago, New York	2008	NGO	local, international
	10	**Social and Behavioral Sciences Team**	USA, Washington DC	2014	public	national
Latin America	11	**Ethos Laboratorio de Políticas Públicas**	Mexico, Mexico City	2008	NGO	national, international
	12	**iGovLab Laboratório de inovação em Governo**	Brazil, San Paulo	2015	NGO of public–academic partnership	national, local
	13	**LabGob Laboratorio de Gobierno**	Chile, Santiago	2014	public	national
	14	**LPP Laboratorio de Políticas Públicas**	Brazil, Rio de Janeiro	2000	academic	national
	15	**Smart Lab**	Argentina, Buenos Aires	2015	public	local
Asia & Australia	16	**Auckland Co-Design Lab**	New Zealand, Auckland	2015	public	regional, local
	17	**NSW Behavioural Insights Unit**	Australia, New South Wales	2012	public	regional
	18	**PS21**	Singapore, Singapore	1995	public	national
	19	**Pulse Lab Jakarta**	Indonesia, Jakarta	2012	NGO–public partnership	national
	20	**Seoul Innovation Bureau**	Korea, Seoul	2013	public	local

Source: Authors' own elaboration on the basis of web content

coherent typologies and comparative frameworks in the current literature, the focus was on creating a primary synthesis and identifying shared characteristics.

Structures: who labs are

The legal status of policy labs substantially varies, with no visible patterns in terms of regional diversity. Half of the selected labs are exclusively owned by public sector entities representing different levels of governance. Strikingly, in the Asia & Australia group there is not a single lab without public sector ownership. Three of the studied labs function as independent non-government organisations (Ethos, Ideas42, Gov42, Kennisland) and another three are run by universities (Design Policy Lab, GovLab, LPP). The remaining four labs take the form of NGOs launched in various inter-organisational partnerships: public–academic (iGov Lab, GovLabPHL), public–NGO (Pulse Lab Jakarta) or between multiple public sector entities (La 27e Region). The labs' focus on public policy can explain this clear dominance of the public sector. Other studies have analysed in more detail how ownership influences labs' activity and characteristics (McGann et al, 2018).

The selected labs operate at local, regional and national levels. Some of them are also active at the international level (for example, Design Policy Lab, Ethos). The majority of labs specialise in one particular level of governance (mostly national), but seven of them operate at more than one level (for example, GoVLab, Alberta CoLab, Auckland Co-Design Lab, iGovLab).

The experience of the studied labs varies from two to 23 years. The oldest five labs in our selection were launched between 1995 and 2002: three European, one Asian and one Latin American. However, it can be highlighted that Mindlab is the oldest institution actually launched under the policy lab name, while others started out as a different organisation and only later transformed into a lab. As a result of the recent boom in lab development (clearly visible in the Americas and Asia), the majority of the studied labs are no more than seven years old. The youngest lab included in the analysis is the Philadelphian GOVLabPHL launched in 2016.

In terms of networks, labs' regular partners usually include local and regional authorities, public administration institutions and private stakeholders, who actively contribute to projects for the common good. They often engage with researchers specialised in a particular policy area. Civil society organisations are an important partner for labs, as they represent wider groups of citizens or specific social interests. As for private sector companies or corporations, they are usually involved as sponsors of projects or as partners in programmes supporting entrepreneurship. Some labs are also associated with various international projects, usually financed by transnational foundations, OECD or the European Commission.

In general, labs diversify policy ownership by engaging various actors or groups representing different stakeholders. Often this function is strengthened by the fact that they operate in a physical space outside an administration with its set routines and standard regulations, which stimulates open discussion among equal partners.

Functions: what labs do

In terms of purpose, the studied labs usually define themselves as platforms or 'shared spaces' of collaboration, knowledge production and implementation, to highlight

their inter-sectoral bridging capacity. Sometimes they also define themselves as action research-oriented units that aim to promote government effectiveness and cultural shifts.

The thematic scope of labs' actual operations varies substantially: some of them focus on a particular sectoral policy (for example, housing and civil society empowerment in the case of Kennisland), while some do not have a particular specialisation, and support policy-design processes depending on their founders' needs. This observation runs in line with previous studies. For example, Australian labs cover the whole spectrum of social policy issues, including housing and welfare, public administration and governance, education, health, transport, policing, as well as the criminal justice system (McGann et al, 2018).

The analysis of policy lab missions and key activities did not reveal any clear geographical patterns. However, across all regions we could identify three main specialisation patterns:

1 serving the government: aiming to improve public services and policies, and train administrative staff (MindLab, UK Policy Lab, PS21, Alberta CoLab, LabGob; iGovLab, GOVLabPHL, La 27e Région). Two labs framed their mission as 'changing the way of governance' (Design Policy Lab, GovLab,);
2 serving the community: aiming to solve specific social problems. Here, labs' areas of interest were diverse and broad, including issues like health, poverty, housing, domestic violence (NSW Behavioural Insights Unit, Seoul Innovation Bureau), youth and families (Auckland Co-design), human rights, urban development (LPP), poverty alleviation, social development, gender equality, democratic governance (ETHOS), making society smarter, empowering citizens (Kennisland);
3 serving both the government and the community (Social and Behavioral Sciences Team, Pulse Lab Jakarta).

In general, policy labs produce programmes and projects that seek to explore ideas, solve problems, train leaders and deliver tools to improve public services through innovation. This includes creating spaces for entrepreneurs, SMEs, students, academics, citizens and NGOs to use their talent, ideas and experiences related to a particular public service in order to find solutions to existing challenges.

Labs aim to create impacts in two major ways: by engaging citizens, and by changing the culture of government through increasing the adoption of suggested changes and improvements by the relevant agencies and departments. However, with regards to the effects of policy labs' activity, the study has found little evidence. It seems that labs' ambition to create meaningful change in these two areas has not been followed by a clear idea of what exact impact they are aiming at, nor a way to measure it. For example, when reporting on 'impact', labs do not discuss long-term, structural change in the situation of the addressees. Instead, they present short-term, mainly process-related outcomes (meetings, reports, and so on). Likewise, their success stories have a highly processual character – they report on connections, networks and actors' involvement, and the scale of dissemination.

This clear mislabelling of process indicators as impact indicators can be partially explained by the short period of existence of some labs, but in other cases it may be associated with a reluctance to make outcomes measurable and public. In other words,

the level of labs' transparency may be influenced by their partners' expectations, as well as the political sensitivity of some matters addressed.

We can conclude that what seems to be lacking is a systemised effort to measure the impact after fully implementing a project. Some labs are satisfied with general feedback from the community, partner institutions or local governments ('Looks like the mayor is happy') often collected during workshops. However, there seems to be little research done in measuring real change in the behaviours of citizens, public servants or organisational actors. The lack of this feedback loop may limit the positive impact of labs' work, as it hinders learning from experience and developing better tools in the future. Therefore, there is a need to develop greater reflexivity in terms of the short- and long-term effects of labs' work.

Processes: how labs do things

The processes and toolbox of policy labs are heavily based on design practice including design thinking and industrial design (Bason, 2014; IDEO, 2015; Stickdorn and Schneider, 2012). Numerous labs refer to various versions of the design process with a spectrum of techniques for each of them (Andrews, 2015; Kieboom et al, 2015; MindLab, 2017; Mortati, 2015; Open Policy Making, 2017). This observation is in line with findings of earlier comparative studies that point out a recurring pattern in service-design rationale, which is a participatory, human-centred approach (McGann et al, 2018).

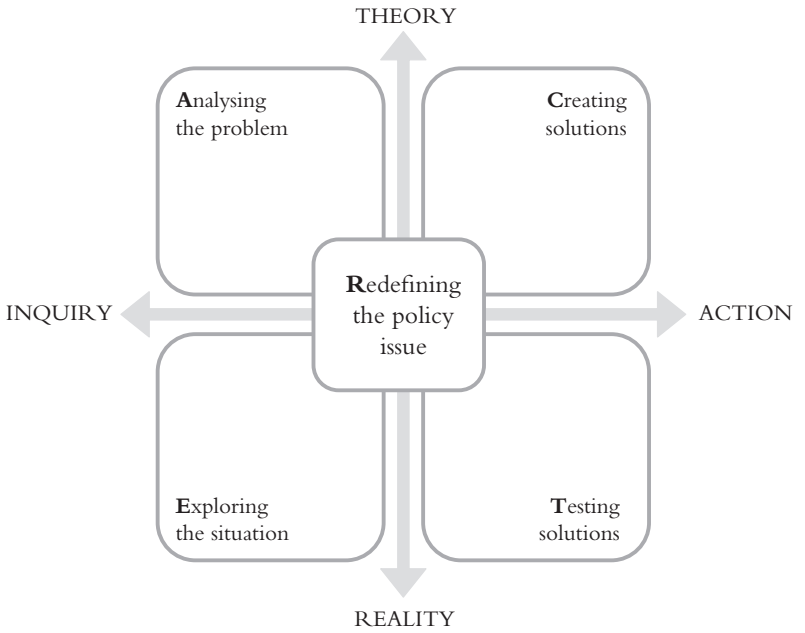
In order to find a broader pattern in lab processes and methods, we decided to map them on a two-dimensional matrix (Kumar, 2012). The first dimension is ontological in nature – it extends from experiential immersion in reality on the one hand, to a more theoretical abstraction of policy issues on the other. It captures well the tension existing in public policy practice, between the complex, often chaotic reality of a particular policy problem, and the conceptual efforts to make sense out of this reality, finding more general principles that could be applied to problem-solving, and even transferred to other policy situations.

The second dimension describes the nature of involvement with a policy issue. It ranges from inquiry – focused on understanding the policy issue, to creation – focused on active manipulation and change of the matters addressed by the particular policy. This second dimension illustrates well the continuous struggle in public policy practice to feed diagnostic studies into decision-making and implementation, or looking more broadly – to integrate the insights from various research inquiries into policy actions.

Mapping the activities of the analysed labs onto this matrix allowed us to uncover five areas of activity accompanied by a variety of tools and techniques (see Figure 1 and Table 3). Following labs' practice of making procedures user-friendly and easy to grasp, we named our generic process REACT. This is an acronym created from the first letters of each of its stages.

First, the work of labs starts with assisting policymakers in redefining a policy issue. The policy problem is often framed in terms of various target groups, segments of policy addressees and their desired behaviours. Second, labs engage in deep exploration of the reality that they want to change. This involves collecting relevant data from multiple sources to fully understand the problems, behaviours and context of target groups. Third, policy labs move into the analysis stage, hypothesising about the factors and contexts that facilitate or hamper the desired policy change. Fourth, they take an

Figure 1: REACT – generic design process in policy labs



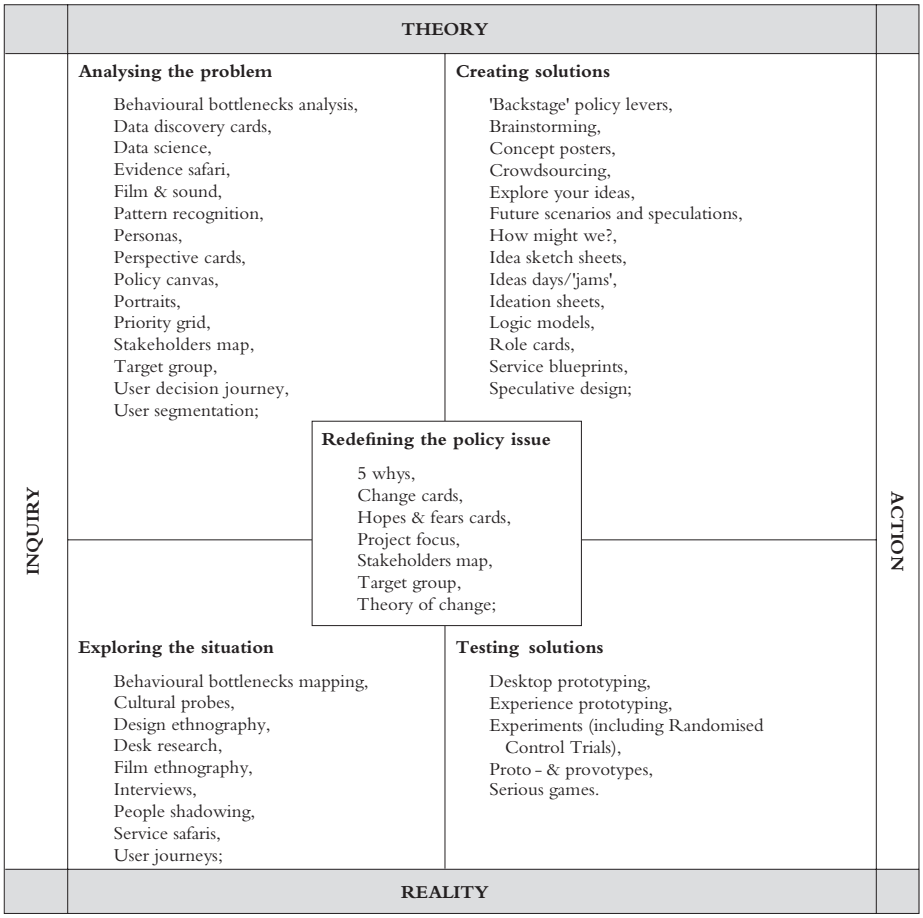
Source: Authors' own elaboration, based on Andrews, 2015; MindLab, 2017; Kumar, 2012 and web content

active role and undertake creative sessions, in which they try to construct potential solutions to the studied problems. Fifth, labs provide a safe space for experimentation, which allows them to test in practice the prototypes of designed solutions. This whole process is often not linear, but iterative. For example, after completing the first series of testing, labs might discover a new group of service users, and therefore decide to further explore the characteristics of this group and repeat the analysis, creation and testing processes.

The REACT matrix (Figure 1) highlights the dialectical nature of applied policy design and the efforts of policy labs to integrate in one stream various activities and perspectives that have been traditionally separated in policymaking. Policy labs constantly balance between abstraction and fine-grained reality, between research inquiry to understand problems and proactive development of solutions to the problems. We believe this is an important characteristic that distinguishes labs from other institutions involved in policy design and evaluation, which traditionally tend to focus on a single sphere (for example, theoretical reflection or implementation).

Looking into the toolbox of policy labs, two patterns seem dominant. First, there is the participatory orientation, focused on tools and methods (for example, ethnographies, interviews, workshops, crowdsourcing, brainstorming) involving representatives of various stakeholders in the exploration and co-creation of solutions. Labs from all continents, including almost all analysed labs from Europe, Australia and Asia, subscribe to this orientation. The human-centred approach welcomes citizens' ideas and solutions, embraces ambiguity and makes it possible to identify the conditions for co-design and implementation. Some labs also highlight the need to involve participants during the prototyping and testing stage, in order

Table 3: Methods and tools used by policy labs



Source: Authors' own elaboration based on Andrews, 2015; MindLab, 2017; Kumar, 2012 and web content

to increase the takeup of solutions, empower beneficiaries and socially embed these solutions.

The second pattern is the use of behavioural insights, characteristic of the Anglo-Saxon labs analysed in this article (GOVLabPHL, Ideas42 & Gov42, Social and Behavioral Sciences Team, Alberta CoLab, NSW Behavioural Insights Unit). These insights are particularly valid for the exploration, creation and testing stages. The first two stages use knowledge about human motivations and decision-making processes. Exploration involves mapping behavioural bottlenecks - factors that obstruct desired behaviours of policy addressees. Creation uses behavioural change strategies (including 'nudging') in the design of potential solutions. The testing stage relies on small, low-cost interventions accompanied by experiments (usually randomised control trials).

The policy labs analysed in this article often highlight that their work aims at scaling-up the tested solutions. However, there is little evidence on how labs get involved in scaling-up and actual policy design. They can lobby, advocate and disseminate their findings among multiple policy actors and stakeholders, but they are usually not directly involved in the full-scale policy design, which some authors

see as one of the biggest limitations of policy labs (McGann et al, 2018). Thus, we adopt the perspective that scaling-up is a separate meta-process that goes beyond the typical cycle of lab activities, and requires a separate set of activities, methods and tools.

Labs and the challenges identified by evaluation practice

The synthesis of lab practices presented in this section makes it possible to explore the ways in which policy labs address or fail to recognise the three challenges identified by evaluation practice.

We start with the **challenge of establishing what solution works**. Looking at current lab practices, it seems that labs have little to contribute to the ongoing debate addressing the challenges of valuing interventions; they could benefit from incorporating the existing body of knowledge from evaluation into their practices.

Labs could learn from evaluators' community discussions that have, over the years, sought to overcome paradigm divides, using more sentient ways of mixing methods and exploring the effects of complex interventions (Stern et al, 2012; Bamberger et al, 2015). Evaluation practice offers a high level of emphasis on obtaining strong causality, namely causality from the successionist perspective, through rigorous evaluation methods such as Randomised Controlled Trials (RCTs). However, recent evaluation trends stress that not only 'what works' is important, but also 'in what context,' and 'for whom' are critical elements in assessing the effectiveness of a policy intervention based on various perspectives and notions of causality. What significantly worked in Setting A might not work in Setting B or C. As generativists of causality argue, the contextual factors or supporting mechanisms of Setting B or C may be different from Setting A. Particularly, the last element 'for whom', leads us directly to the issue of assessing the value of evaluation findings, since it involves making judgements for the beneficiaries of the evaluation. This implies that the evaluation inquiry requires evolving practices that balance rigour and flexibility in the use of methods and analytical approaches. Moreover, creating synergetic effects between the use of labs and evaluation can promote making value judgements of the findings. This can be achieved by incorporating stakeholder perspectives and different value judgements at the early stage of the programme design and by addressing the needs and interests of programme beneficiaries, in particular, enhancing the responsiveness of previously underrepresented populations in evaluation.

In case of the second issue, **explaining why a solution works or not**, there are three promising synergies emerging from labs and evaluation practice that could address the challenge of unpacking the black box of mechanisms.

Both evaluation practice and labs use a theory-driven approach that defines public intervention as a 'theory' – a hypothesis yet to be tested and verified in a real-life situation (Rossi et al, 1999; Chen, 2004; Coryn et al, 2011; Donaldson, 2007). The real value of the theory-driven approach, especially for decision-makers, is its explanatory dimension. It reveals how a programme works, with whom and under what circumstances (Astbury and Leeuw, 2010: 365), and explains the mechanisms that have led to the success or failure of the intervention.

The five-stage REACT process applied by labs (compare Figure 1) seems to provide Programme Theory building that is easier than the reverse process of theory building in evaluation. In labs, the exploratory and analytical phases help articulate the 'hypothesis on obstructing problems' (HOP) – the key barriers that block the

desired behaviour. Then, the creative phase provides ideas for solutions – that is, the 'hypothesis on intervention types' (HITs) that can address these barriers. When the testing phase shows limited results, designers can go back and reconsider either their diagnosis of the main barrier (that is, HOP) or the selection of policy instruments (HIT). This logic is aligned with the human cognitive process of everyday hypothesis testing as part of problem solving (Evans, 2017). As a result, it is more intuitive and accessible to practitioners.

Second, labs add an important aspect to the explanation of mechanisms – behavioural insights into the users on a particular public intervention. The basic strategy here is to break down the 'journey' of the policy actor throughout the intervention into a sequence of interactions and decisions, and then analyse 'behavioural bottlenecks', including individual heuristics and biases that can block desired behaviours (Ly et al, 2013; Service et al, 2015; Stephans, 2016). This provides a number of advantages: it allows exploration of observed (not declared) behaviours and their drivers, clarifies the synthesis of diagnostic findings in the form of user profiles and their decision journeys, and links policy practice with the growing body of psychological and cognitive research on human decision-making. This approach also focuses the attention of decision-makers on concrete change in behaviours, and allows the formation of hypotheses on the barriers or gaps that hinder such change.

Finally, labs combine approaches in constructing intervention theory that are usually separated in evaluation practice. They use a deductive approach by referring to scholarly theories (mainly behavioural science and design) and earlier intervention results from academic literature. They also use an inductive approach – conducting fieldwork and generating a grounded theory. Labs add to these a user-focused approach – working with intended users to extract and make explicit their implicit underlying assumptions as well as their theory of change. Leading evaluation theorists have been calling for such a combined approach (Patton, 2010; Donaldson, 2007).

The last **challenge of transferring research findings into policy actions** has been the focus of both the theory and practice of evaluation. This body of knowledge, when combined with lab practices could lead to promising developments in ameliorating the problem of evidence-informed policy.

Labs' scope of operations provides perspectives for better utilisation of research by addressing the issue of timing. The majority of labs focus on working at the early stages of the policymaking process – that is, problem definition and policy formulation. Some labs (for example, GovLab) focus on improving already existing services. What is especially interesting is that labs often try to build, within the main policy cycle, a smaller loop of design–testing–adaptation (Haynes et al, 2012). They gain some time by applying small-scale experiments and testing prototypes. This allows them to deliver meaningful feedback on 'what works and why' exactly when decision-makers need it – relatively early on in the policy process.

Both evaluation practices and labs are well aware of the importance of overcoming the 'language' barrier between researchers and decision-makers. The evaluation tradition has focused on improving ways and forms of communication with potential users of knowledge. This includes studies on understanding the credibility of information in the eyes of the beholder (Miller, 2014), effective data visualisation (Azzam et al, 2013) and broader knowledge brokering strategies (Olejniczak et al, 2016). To this, labs could add their organisational structures that facilitate the utilisation of their products. The involvement of stakeholders and partnership structures with

policy actors would build engagement and a sense of ownership of the developed solutions. This, combined with co-designing initiatives, could address the gap between producers and users of knowledge, since all interested parties would be involved in the process.

Conclusions

The reviewed practices of labs seem indeed promising in addressing the challenges of public policy. However, the methodological limitations of our current study (with analysis based only on desk research of web sources and labs' own reports) mean that at least three issues would benefit from further research.

First, our study was aimed at addressing the current lack of coherent typologies and comparative frameworks among policy labs. This focus on synthesis and shared characteristics has limited the discussion of heterogeneity among labs. Future comparative studies could use the analytical frames developed in this article (typologies of labs, generic matrix of the design process) to explore geographical and topical variations (including the growing phenomenon of digital policy labs), and factors influencing the diversity of labs.

Second, it is unclear how effective the analysed labs are in feeding their solutions into the actual policymaking and policy implementation process. Are they stand-alone initiatives that strive for decision-makers' attention, and occasionally feed ideas into the policy cycle, or are they integrated into the institutional policy system of a particular region or government? These questions have been difficult to address through this research due to the short history of lab operations. A new cycle of interviews with lab personnel could provide us with better understanding of how systematic their role is in the policy process.

This leads to the last limitation – we do not know how sustainable the labs are in the long run and what make them durable in organisational terms. Based on current observations, we can hypothesise that labs set up in a local or regional government by a single public policy leader will have a lower survival rate than initiatives established by a coalition of academic researchers and government representatives. However, this hypothesis can only be verified in the future. With the majority of the labs being relatively young (2–3 years old), it is still far too early to decide if they are just a temporary fashion or an approach that will substantially change the way we design our public policies.

Despite these limitations, through our analysis of the practices of policy labs around the world, we have identified a number of synergies between policy labs and evaluation that could be beneficial for both, and that could advance general public policy practice.

In relation to the challenge of establishing what works and valuing public interventions, policy labs are more likely to be beneficiaries than being contributors in advancing current practices. Policy labs focus on services and projects that are relatively simple and manageable interventions, rarely developing policy or even programme solutions. They have a tendency to use methods that infer relatively quick, short-term and strong causal relationships. However, public programmes have grown in complexity (multi-level, multi-objective interventions) often with extended timeframes (multi-annual planning). Furthermore, today's policy issues are often wicked policy problems that require a more sophisticated understanding of the social dynamics and environments surrounding a certain policy intervention.

Thus, well-established evaluation practices can support labs in tackling issues, ranging from conceptual and processual complexity to the non-linear causality of more complex interventions. In particular, evaluation brings together various perspectives of causality and complexity theories that can help labs measure their impact and understand how systems change over time as well as from place to place (Bamberger et al, 2015; Patton, 2010; Pawson, 2013).

Additionally, insights from evaluation theories and practices can also help policy labs improve the reliability and value assessment of their prototype solutions. There is a range of evaluative criteria suggested by evaluation practitioners, and therefore results may vary depending on the criteria applied. Moreover, the available timeframe changes the results of assessments. Policy labs currently tend to focus on generating short-term effects by using RCTs, anticipated declared utility or satisfaction tests by stakeholders. In the future, policy labs should identify and apply more advanced approaches to valuing.

With regards to the second challenge of explaining change mechanisms, labs and evaluation share common ground in defining public interventions as 'theories' yet to be tested and verified in real-life situations. Lab practices can add to evaluation practice their strong focus on the final users of policy (policy addressees) – their identification, understanding and targeting. Some labs seem to be ahead of evaluation practice in terms of integrating behavioural insights into public policy. Furthermore, the generic REACT process offered by labs to constructing intervention theory is better aligned with human problem solving, and is therefore more likely to be accommodated by policy practitioners.

Looking at the third challenge of feeding research findings into decision-making, policy labs seem to be better equipped than evaluation in addressing the timing issue of delivering results to knowledge users. Labs can help to determine the feasibility and scalability of policy programmes in a relatively flexible timeframe, through means such as implementing ex-ante pilot projects.

Policy labs also offer collaborative structures. They are used as a safe space that encourages the involvement of citizens and stakeholders in co-designing public solutions. These organisational arrangements can help to serve the 'use' purpose of evaluation. They can be important in the policy feedback loop process. This is particularly significant because labs place the tools directly in the hands of the decision-makers, encourage them to partner up with researchers to generate relevant data and evidence, and make assessments to promote the effectiveness of policies and programmes.

Evaluation theories can guide policy labs to think more about their systemic influence and their sustainability. Labs declare an ambitious aim of cultural shift, but they seem to have limited recognition of the complex mechanisms driving organisational learning, including inter-organisational collaborations and political discourses. Evaluation theorists such as Carol Weiss and Peter Rossi offer effective frameworks and principles for those issues. Evaluation practice can also contribute to scaling-up, by emphasising the need for external validity, using approaches such as experiments that are larger in scale in terms of the greater representativeness of samples or increased implementation scale.

To conclude, policy labs are not an alternative development to established practices of public policy. They are rather a promising addition that, when combined with existing evaluation practices, can allow both entities to mutually benefit one another and create synergies to improve the craft of public policy.

Funding details

This work was supported by Narodowe Centrum Nauki under Grant number 2014/13/B/HS5/03610.

Open access of this article was financed by the Ministry of Science and Higher Education in Poland under the 2019-2022 program "Regional Initiative of Excellence", project number 012 / RID / 2018/19.

Acknowledgements

The authors would like to thank the team of editors of the special issue, the editors of the journal and the two anonymous reviewers for their most valuable comments that substantially improved the content of this article.

Conflict of interest

The authors declare that there is no conflict of interest.

References

Alkin, M.C. (2013) *Evaluation roots: Tracing theorists' views and influences*, Thousand Oaks, CA: Sage.

Amadae, S. (2007) Rational choice theory, In M. Bevir (ed), *Encyclopedia of governance*, Thousand Oaks, CA: Sage, pp 785–91.

Andrews, B. (2015) *Policy lab: Method bank and toolkit*, London: Policy Lab UK, www.slideshare.net/Openpolicymaking/methodbank-and-toolkit-for-design-in-government

Astbury, B. and Leeuw, F.L. (2010) Unpacking black boxes: Mechanisms and theory building in evaluation, *American Journal of Evaluation*, 31(3): 363–81. doi: 10.1177/1098214010371972

Azzam, T., Evergreen, S., Germuth, A. and Kistler, S. (2013) Data visualization and evaluation, *New Directions for Evaluation*, 139(Fall): 7–32. doi: 10.1002/ev.20065

Bamberger, J.J.M., Vaessen, J.L. and Raimondo, E.R. (2015) *Dealing with complexity in development evaluation: A practical approach*, Thousands Oaks, CA: Sage.

Bason, C. (2014) *Design for policy*, Farnham: Gower.

Batterbury, S. (2006) Principles and purposes of European Union cohesion policy evaluation, *Regional Studies*, 40(2): 179–88. doi: 10.1080/00343400600600504

Bohni Nielsen, S., Lemire, S. and Christie, Ch. (2018) The evaluation marketplace: exploring the evaluation industry, *New Directions for Evaluation*, (160) Winter: 1–163.

Caplan, N. (1979) The two-communities theory and knowledge utilization, *American Behavioral Scientist*, 22(3): 459–70. doi: 10.1177/000276427902200308

Chelimsky, E. (1985) *Program evaluation: Patterns and directions*, Washington, DC: American Society for Public Administration.

Chelimsky, E. (1997) The coming transformation in evaluation, In E. Chelimsky and W. R. Shadish (eds) *Evaluation for the 21st century: A handbook*, Thousand Oaks, CA and London: Sage, pp 1–26.

Chen, H.T. (2004) *Practical program evaluation: Assessing and improving planning, implementation, and effectiveness*, Thousand Oaks, CA: Sage.

Colebatch, H.K. and Hoppe, R. (ed.) (2018) *Handbook on policy, process and governing*, Cheltenham: Edward Elgar.

Coryn, C., Noakes, L., Westine, C. and Schroter, D. (2011) A systematic review of theory-driven evaluation practice from 1990 to 2009, *American Journal of Evaluation*, 32(2): 199–226. doi: 10.1177/1098214010389321

Davies, H., Nutley, S. and Smith, P. (eds.) (2009) *What works? Evidence-based policy and practice in public services*, Bristol: Policy Press.

Derlien, H.U. (1990) Genesis and structure of evaluation efforts in comparative perspective, *Program evaluation and the management of government: patterns and prospects across eight nations*, New Brunswick, NJ: Transaction Publishers.

Donaldson, S.I. (2007) *Program theory-driven evaluation science: strategies and applications*, New York: Lawrence Erlbaum.

Dunn, W.N. (2017) *Public policy analysis: An integrated approach*, New York: Routledge.

European Commission (1995) *MEANS handbook: Volumes 1–7*, Brussels: Commission of the European Communities.

Evans, J. (2017) *Thinking and reasoning: A very short introduction*, Oxford: Oxford University Press.

Fuller, M. and Lochard, A. (2016) *Public policy labs in European union member states*, Luxemburg: Publications Office of the European Union.

Furubo, J.E., Rist, R.C. and Sandahl, R. (2002) *International atlas of evaluation*, New Brunswick, NJ: Transaction Publishers.

Haynes, L., Service, O. and Goldacre, B. (2012) *Test, learn, adapt: Developing public policy with randomised controlled trials*, London: Cabinet Office - Behavioural Insights Team.

Henry, G.T., Mark, M.M. (2003) Beyond use: understanding evaluation's influence on attitudes and actions, *American Journal of Evaluation*, 24(3): 293–314.

Howlett, M. (2011) *Designing public policies: Principles and instruments*, London and New York: Routledge.

IDEO (2015) *The field guide to human-centered design*, Stanford, CA: IDEO.org.

Jacob, S., Speer, S. and Furubo, J.E. (2015) The institutionalization of evaluation matters: updating the international atlas of evaluation 10 years later, *Evaluation*, 1(21): 6–31.

Kahneman, D. (2011) *Thinking, fast and slow*, New York: Farrar, Straus and Giroux.

Kieboom, M., Sigaloff, C. and van Exel, T. (2015) *Lab Practice: Creating Spaces for Social Change*, Amsterdam: Kennisland.

Kumar, V. (2012) *101 design methods: A structured approach for driving innovation in your organization*, Hoboken, NJ: Wiley.

Lasswell, H.D. (1951) The policy orientation, In D. Lerner and H.D. Lasswell (eds) *The policy sciences: Recent developments in scope and method*, Stanford, CA: Stanford University Press.

Leeuw, F.L., Rist, R.C. and Sonnichsen, R.C. (1994) *Can governments learn? Comparative perspectives on evaluation and organizational learning*, New Brunswick, NJ and London: Transaction Publishers.

Ly, K., Mazar, N., Zhao, M. and Soman, D. (2013) *A practitioner's guide to nudging*, Toronto: Rotman School of Management, University of Toronto.

McGann, M., Blomkamp, E. and Lewis, J. (2018) The rise of public sector innovation labs: experiments in design thinking for policy, *Policy Sciences*, 51(3): 249–67. doi: 10.1007/s11077-018-9315-7

Mark, M.M., Henry, G.T. and Julnes, G. (2000) *Evaluation: An integrated framework for understanding, guiding, and improving policies and programs*, San Francisco, CA: Jossey-Bass.

Mathison, S. (ed.) (2005) *Encyclopedia of evaluation*, Thousand Oaks, CA: Sage.

Mertens, D.M. (2008) *Transformative research and evaluation*, New York: Guilford Press.
Miller, R. (2014) How people judge the credibility of information: lessons for evaluation from cognitive and information sciences, In S. Donaldson, C. Christie and M. Mark (eds), *Credible and actionable evidence: the foundation for rigorous and influential evaluations*, Thousand Oaks, CA: Sage, pp 39–61.
MindLab (2017) *Method Guide*, Copenhagen: MindLab.
Mortati, M. (2015) A framework for design innovation: present and future discussions, *Design Issues*, 31(4): 4–16, doi: 10.1162/DESI_a_00347
Newcomer, K.E., Hatry, H.P. and Wholey, J.S. (eds.) (2015) *Handbook of practical program evaluation*, San Francisco, CA: Jossey-Bass.
Nutley, S.M., Walter, I. and Davies, H.T.O. (2007) *Using evidence: How research can inform public services*, Bristol: Policy Press.
OECD (Organization for Economic Cooperation and Development) (1998) *Best practice guidelines for evaluation*, Paris: OECD.
Olejniczak, K. and Mazur, S. (eds.) (2014) *Organizational learning: A framework for public administration*, Warsaw: Scholar Publishing House.
Olejniczak, K., Raimondo, E. and Kupiec, T. (2016) Evaluation units as knowledge brokers: testing and calibrating an innovative framework, *Evaluation*, 22(2): 168–89. doi: 10.1177/1356389016638752
O'Rafferty, S. (2016) *Labs, iTeams, designers*, Limerick: Co-creating Ireland. https://cocreatingireland.wordpress.com/2016/03/29/policy-labs/
Open Policy Making (2017) *Open policy making toolkit*, www.gov.uk/guidance/open-policy-making-toolkit
Patton, M.Q. (2004) Evaluation research, In M. Lewis-Beck, A. Bryman and T. Liao, (eds), *The sage encyclopedia of social science research methods*, eds. M. Lewis-Beck, A. Bryman, T. Liao (Thousand Oaks, CA: Sage), pp 337–40.
Patton, M.Q. (2010) *Developmental evaluation: Applying complexity concepts to enhance innovation and use*, New York: Guilford Press.
Pawson, R. (2006) *Evidence-based policy: A realist perspective*, London: Sage.
Pawson, R. (2013) *The science of evaluation: A realist manifesto*, London: Sage.
Pawson, R. and Tilley, N. (1997) *Realistic evaluation*, London: Sage.
Price, A. (2015) *World of labs*, London: Nesta, www.nesta.org.uk/blog/world-labs
Rossi, P.H., Freeman, H.E. and Lipsey, M.W. (1999) *Evaluation: A systematic approach* (6th edn.), Thousand Oaks, CA: Sage.
Scriven, M. (1977) *Evaluation thesaurus*, Newbury Park, CA and London: Sage.
Service, O., Hallsworth, M., Halpern, D., Algate, F., Gallagher, R., Nguyen, S. and Sanders, M. (2015) *EAST: Four simple ways to apply behavioural insights*, London: Behavioural Insights Team.
Shafir, E. (2013) *The behavioral foundations of public policy*, Princeton, NJ: Princeton University Press.
Shaw, I., Greene, J.C. and Mark, M.M. (eds.) (2006) *The SAGE handbook of evaluation*, Thousand Oaks, CA: Sage.
Shillabeer, A., Buss, T.F. and Rousseau, D.M. (eds.) (2011) *Evidence-based public management: Practices, issues, and prospects*, Armonk, NY: ME Sharpe.
Simon, H.A. (1997) *Administrative behavior* (4th edn.), New York: The Free Press.
Stephans, E. (2016) *Social and behavioral science used in the advancement of policy: applications and insights*, New York: Nova Science Pub Inc.

Stern, E., Stame, N., Mayne, J., Forss, K., Davies, R. and Befani, B. (2012) Broadening the range of designs and methods for impact evaluations, *Working Paper 38* Washington, DC: Department of International Development.

Stickdorn, M. and Schneider, J. (2012) *This is service design thinking: Basics, tools, cases*, Hoboken, NJ: Wiley.

Tõnurist, P., Kattel, R. and Lember V. (2017) Innovation labs in the public sector: what they are and what they do?, *Public Management Review*, 19(10): 1455–79.

United Nations Evaluation Group (2013) *International year of evaluation 2015*, www.unevaluation.org/document/download/2145

Weaver, K. (2015) Getting people to behave: research lessons for policy makers, *Public Administration Review*, 75(6): 806–16. doi: 10.1111/puar.12412

Weiss, C.H. (1988) Evaluation for decisions: is anybody there? Does anybody care?, *Evaluation Practice*, 9(1): 5–19. doi: 10.1016/S0886-1633(88)80017-5

Weiss, C.H. (1999) The interface between evaluation and public policy, *Evaluation*, 5(4): 468–86. doi: 10.1177/135638909900500408

Wildawsky, A. (2017) *The art and craft of policy analysis*, Berkeley, CA: Palgrave Macmillan.

World Bank (2015) *Mind, society, and behavior*, Washington D.C.: World Bank Group.