



Accepted for publication 12 March 2019 • First published online 22 July 2019
This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 license (<http://creativecommons.org/licenses/by/4.0/>) which permits adaptation, alteration, reproduction and distribution for non-commercial use, without further permission provided the original work is attributed. The derivative works do not need to be licensed on the same terms

SPECIAL ISSUE • Co-creative approaches to knowledge production and implementation

practice

An experimental evaluation tool for the Public Innovation Lab of the Uruguayan government

Cristina Zurbriggen, criszurbriggen@gmail.com

University of the Republic, Uruguay and South American Institute for Resilience and Sustainability Studies, Maldonado, Uruguay

Mariana González Lago, mariana.gonzalezlago@gmail.com

University of the Republic, Uruguay and University of Technology, Sydney, Australia

In many parts of the world, governments are building new platforms, methods, and innovative experimental spaces to better respond to current complex problems. Laboratories in the public sector have emerged as experimental spaces that incorporate co-creation approaches to promote public innovation and social transformation. Although there is abundant literature about public innovation and reports on innovative practices, little progress has been made on how to evaluate these. In this paper, we describe the process that led to the development of an experimental evaluation tool for public innovation as part of an action-research process in a laboratory within the Uruguayan Government. The pilot prototype, the 'Roadmap' as we named it, seeks to provide a timely and purposeful means to learn from the co-creation processes and be accountable to public authorities and society. Aiming to build a learning system within the organisation to communicate results, we designed the Roadmap based on the confluence of various approaches, namely, development evaluation, organisational learning and reflexive monitoring. Other relevant approaches to public innovation and evaluation were also considered, such as public design evaluative thinking, social innovation evaluation, and systemic evaluation of learning.

key words public innovation • laboratory • co-creation • evaluation

Key messages

- We designed an evaluation and monitoring tool to promote social learning in co-creation approaches for the design of public services and policies.
- The prototype seeks to provide a timely and purposeful means to learn from co-creation and be accountable to public authorities and society.
- It is a heuristic tool rather than a set of prescriptive instructions about how to evaluate public innovation.

To cite this article: Zurbriggen, C. and González Lago, M. (2019) An experimental evaluation tool for the Public Innovation Lab of the Uruguayan government, *Evidence & Policy*, vol 15, no 3, 437–451, DOI: 10.1332/174426419X15537488717501

Introduction

Innovation in public policies and services is moving to the top of the agenda at all government levels in many parts of the world. Governments are exploring different ways to involve public servants, the private sector, civil society and academia to play an active role in tackling complex problems (Agger and Lund, 2017). In recent years, a multiplicity of experimental spaces – that is, the laboratories, commonly known as ‘Labs’ – have emerged to promote public innovation and social transformation, incorporating co-creation approaches for the design of public services and policies (Peters and Rava, 2017). The number of innovation Labs in government has rapidly grown in the last decade (Bekkers et al, 2011; Ansell and Torfing, 2014; Agger and Sørensen, 2014; Timeus and Gascó, 2018; Tönurist et al, 2017; Puttick et al, 2014; Mulgan, 2014; Bason, 2010).

Labs are conceived as ‘islands of experimentation’ (Tönurist et al, 2017, 8) where creative methods are used ‘to change the way government operates’ (Bason and Schneider, 2014, 35), involving ‘all stakeholders in the process’ (Fuller and Lochard, 2016, 1). In particular, public sector innovation Labs use experiment-oriented approaches to policy and service design to address the systemic nature of policy and social challenges (Fuller and Lochard, 2016; Ansell and Bartenberger, 2016; McGann, Blomkamp, and Lewis, 2018; Kimbell, 2016; Junginger 2017).

Two Labs born in Western Europe, the MindLab (Denmark) and the Behavioural Insights Team (UK), have been the source of inspiration for governments across the world to re-imagine their public services and create similar initiatives. Some examples include the Seoul Innovation Bureau in South Korea, the Centre for Public Service Innovation in South Africa, the Public Innovation Lab in Chile, the OPM Innovation Lab in Washington D.C., the Co-Lab in Sweden, ARENA A-Lab in Australia, the Laboratory for the City in Mexico City, The Human Experience Lab – THE lab – in Singapore, among many others. Also, different networks have emerged, such as the EU Policy Lab; UNDP and the Sustainable Development Goals-Lab (UNLEASH), to expand this new conception of public innovation (OECD, 2018).

Influenced by the work of the MindLab, the Uruguayan government launched the ‘Social Innovation Laboratory for Digital Government’ in 2015, within the National Agency of Electronic Government and the Information and Knowledge Society (AGESIC by its acronym in Spanish). The purpose of the Lab is to build a culture of innovation oriented towards the democratisation of public management, based on new paradigms of public intervention that emphasise the active participation of citizens in the construction of services and policies. To this end, the Lab’s strategy is to orchestrate experimental and creative processes to understand, empathise with, and devise solutions to current challenges in digital government.

The Lab, like many other similar laboratories, faced adversity and frustration when reporting its results to authorities within the organisation, because the rich information emerging from the co-creation processes hardly conforms to the requirements of the dominant instrumental model of evaluation in the public sector. Although involving citizens and other end users in collectively framing problems and ideating solutions may be an important normative ideal, there is little evidence that demonstrates whether this produces better policies and public service innovations (Voorberg et al, 2015, 1341). Moreover, evaluation in this area lacks a theoretical framework (Dayson, 2016;

Graddy-Reed and Feldman, 2015; Bund et al, 2015), and there are few empirical experiences in these types of public innovation spaces (OECD, 2017).

The Lab, therefore, required a different approach to evaluation rather than using the instrumental model that assumes change can be planned, based on rational problem-solving procedures and using predictive expert knowledge about causes and effects in human behaviour as well as in societal dynamics. This paper will focus on describing the process that led to the development of an experimental evaluation tool for public innovation as part of an action-research process in the Lab. The pilot prototype, the 'Roadmap' as we named it, seeks to provide a timely and purposeful means to learn from the co-creation processes to communicate results and, thus, be accountable to public authorities and society. To build this tool we draw primarily from the guides and principles of developmental evaluation (Westley, Zimmerman, and Patton, 2006; Patton, 2010; 2011; 2015), organisational learning (Argyris and Schön, 1978; 1996; Stringer, 2007), and reflexive monitoring (Arkesteijn et al, 2015; Van Mierlo et al, 2010; 2013). Other relevant approaches to public innovation and evaluation were also considered, such as public design evaluative thinking (Bason, 2010), social innovation evaluation (Dayson, 2016) (Antadze and Westley, 2012), and systemic evaluation of learning (Midgley, 2003; Miyaguchi and Uitto, 2017).

This paper is structured in the following way: the first section introduces the case background, explaining the challenges of evaluation in public innovation, and refers to the critical literature on the topic as the first step to building the conceptual background for the Roadmap. The second section presents the action-research process that led to the development of the Roadmap as an experimental evaluation tool suitable to the specificities of the Lab. Lastly, we present a discussion and conclusions about the evaluation of public innovation, based on the experience of this experimental practice-context in the Uruguayan government.

Background

Challenges in public innovation evaluation

In the last decade, rich literature about public innovation emerged, framing it as a process of co-creation with citizens having an active role to identify solutions to public problems (Blomqvist and Levy, 2006; Hartley et al, 2013; Ansell and Torfing, 2014). One of the most prominent practitioners in the field of public innovation, Christian Bason (2010, 8), defines it as 'the process of creating new ideas and converting them into value for society'. The driver of change under this public innovation lens is, thus, the creative and experimental process that involves diverse stakeholders in generating knowledge collaboratively in the production of new services and public policies and, thereby, becoming the locus of value creation (Agger and Lund, 2017).

Nevertheless, according to McGann and colleagues (2018, 215), the concept of public innovation is not new, and its origins can be traced back to the 1980s when the New Public Management (NPM) discourse of 'reinventing government' emerged. Furthermore, the authors state that such discourse was followed by the creation of several public innovation Labs seeking to make governments more efficient. Public innovation is narrowly defined in NPM as new or improved services and policies. Under this prevailing conception, public innovation follows the rationale of policy making as a process of intelligence-design-choice, in

which public servants ‘apply forethought to guide organizational action to solve problems’ (Bason, 2014, 229).

Contemporary approaches to public innovation distance themselves from NPM because, instead of resorting to scientific evidence-based policy making, they are grounded on a strong experimental orientation to policy and service design. Bason (2014) states that these emerging experiment-oriented approaches to policy design disclose the ‘sensemaking policymaker’ who practises design-intelligence-choice ‘by paying closer attention to how problems are represented’. As Bason (2014, 299) puts it, ‘Design becomes the shaping of things while engaging with others in the flow of action and the production of outcomes.’

Labs provide the necessary ‘room’ to develop these new ways of doing things in government, allowing the experimental paradigm to unfold. Labs normally perform processes of ‘generative’ experimentation (Ansell and Bartenberger, 2016), where a solution concept (an idea, design, program, project, and so on) to a particular problem is created, and iteratively refined based on continuous feedback from the stakeholders immersed in the experiment. A key characteristic of any innovation initiative in these scenarios is that it is rarely clear how or if it will lead to a specific result at all, because of these multiple interactions and potential conflicts arising from values and perceptions in dispute (Patton, 2011).

As has already been recognised by Van de Poel, Mehos, and Asveld (2017), a political challenge that generative experiments face results from the tension ‘between the goals of learning and demonstrating success’. As the authors further argue, this sort of experiment is pressed to show results, enduring the pressures between the aims of evaluating an idea, exploring its limits and demonstrating that it works. For this reason, when a public innovation experiment is subjected to the dominant instrumental model of evaluation that focuses on outputs rather than outcomes, emergent learning processes are neglected and the data obtained can be biased in an effort to produce measurable results, which can ultimately affect the innovation process itself (Morris, 2011).

The emergence of the Lab and the tensions rising from demonstrating results

Uruguay has turned into a regional and global reference in digital government, becoming the first Latin American country to be part of the most advanced digital nations network, the Digital 9 (D9) in 2018. In this context, the Social Innovation Laboratory for Digital Government (in short named the Lab), was created with the purpose of supporting Uruguay’s digital government strategy by experimenting with the uses and applications of digital technologies in people’s daily lives, applying co-creation methodologies. It seeks to promote and disseminate the principles of public and social innovation to improve internal processes, policies and public services. In this way, the Lab contributes to the construction of a culture of creativity and open collaboration in the AGESIC as well as other sectors within the public administration, involving citizen participation.

The Lab developed an experimental strategy of intervention which is adaptive to each project and unfolds in four phases: understanding, empathising, devising and experimenting. The strategy aims to innovate in tools and techniques based on different design approaches (for example, design thinking, human-centred design, usability, accessibility of users, agile methodologies, games, among others) and harnessing

the team's diverse disciplinary backgrounds in Anthropology, Social Psychology, Ethnography, Communication, Design, and Engineering.

Until now the Lab has assisted in the digitisation of administrative procedures across all sectors in government, incorporating diverse stakeholders' insights and experiences to provide better user-friendly solutions, as well as optimise time and reduce paper use. As a result of the project 'Online Procedures', the Lab has run 39 co-creation workshops to redesign these services with the participation of 154 public servants and 83 citizens; 50 prototypes were created, and today 33 new procedures are online.

However, despite the support it has received from inside and outside the government, it has found 'barriers' in its institutional insertion due to traditional public organisations' management idiosyncrasy: adversity to risk and limited tolerance to failure, resistance to change, and bureaucratic/administrative routines. Consequently, it became strongly dependent on political support to thrive in the face of cultural resistance and the lack of a shared language.

Public innovation labs often confront the risk of being isolated from their parent organisation, which limits their overall impact on innovation capacity and questions the sustainability of innovation in the public sector (Timeus and Gascó, 2018). The Lab is not an exception to this. Since its creation, it has been tackling the challenge of legitimising itself and demonstrating (successful) results within AGESIC, as well as to other national authorities and financing international organisations.

Given the strain between the difficulties of conforming to the requirements of instrumental evaluation and the growing concern to better reflect their outcomes, in 2017, after two years of operation, the Lab enlisted a team from the University of the Republic to work with them on designing an evaluation tool tailored to the organisation. We proposed to develop a pilot, a Roadmap, as we named it, to guide the Lab in harnessing learnings from co-creation processes and communicating public innovation outcomes to government authorities and the broader audience (Hellstrom 2013; 2015). The project was conducted from February 2017 to February 2018, and was based on an action-research approach, focusing on social learning and context adaptation (Bammer, 2005; 2017; Pohl et al, 2008; Klein, 2008).

To develop the Roadmap, we were embedded in the Lab routines for a year. The process of research included fieldnotes and photographs; documents produced with the Lab such as presentations, project briefs, reports, summaries of meetings and emails. In addition, we conducted participant observation, semi-structured interviews with the Lab's team (6), civil servants (9), and stakeholders (10), who participated in the Lab's co-creation workshops, as well as with the directors of three Latin American Labs (Chile, Argentina and Colombia). The Roadmap was designed through iterative cycles of literature review, identifying relevant themes in the evaluation process, sharing them with the Lab team, triangulating with information arising from interviews, and referring back to scholarly and grey literature.

The conceptual framework behind the Roadmap

In order to design a purposeful evaluation and monitoring tool for the Lab, we started by exploring new conceptual frameworks about innovation evaluation. Scholars and practitioners have been in a constant search for the operationalisation of innovation. A tipping point in that quest was initiated by Milbergs and Vonortas (2004) with the introduction of the 'fourth generation of innovation metrics'. The authors expressed, in this way, the necessity to move from a linear conception of innovation rooted in

the industrial economy, to a systemic, non-linear interpretation focused on flows and processes and, thus, better adapted to a knowledge-based economy. For social innovation scholars [Antadze and Westley \(2012\)](#), the proposal of Milbergs and Vonortas was a critical contribution to the analysis of innovation metrics, recognising it as a multidimensional, uncertain and unpredictable process.

However, the most significant change to advance in the evaluation of public innovation occurred in 2006 when Westley, Zimmerman and Patton published: 'Getting to Maybe: How the World Is Changed', that served to set the philosophical bases of what later on will be called 'Developmental Evaluation' (DE). In the years following the publication, the guiding principles for this model of evaluation were consolidated ([Gamble, 2008](#); [Patton, 2010](#); [Dozois et al, 2010](#); [Preskill and Beer, 2012](#)), giving birth to a new evaluation paradigm oriented to learning and adaptation in complex systems ([Snowden and Boone, 2007](#)).

[Patton \(2011\)](#) defines DE as an evaluation that informs and supports innovative and adaptive interventions in complex dynamic environments, in real time. This model seeks to achieve changes in the way of thinking and behaviour of the stakeholders, and in the procedures and organisational culture resulting from the learning generated during the evaluation (ontological). DE is based on the American pragmatist tradition ([Dewey, 1927](#)) using qualitative and quantitative methodologies in the process of knowledge construction (epistemology). The model involves the stakeholders who should review how the evaluation would contribute to developing a shared vision regarding the intervention (programme or project); how it can support and reinforce the intervention, strengthen skills, knowledge, and appropriation of the people involved; and what effects the measurements made can have in the organisational dynamics ([Patton et al, 2013](#)).

In the DE the unit of analysis is no longer the project or programme but the system. Instrumental models of evaluation (that is, summative and formative) were created and developed within the boundaries of projects, and for that reason have a strong project-based mentality ([Scriven, 1996](#)). However, in the last 40 years there has been a growing recognition that projects alone do not lead to change. Change is sustainable when it is systemic, and projects are only a small part of the big picture. Therefore, if a project aims to obtain certain results, it has to affect the system where it is embedded, which at the same time is open, dynamic, and with multiple intervening factors not controllable by the intervention.

A key idea of this post-positivist model is that evaluation should help guide collaborative action and strategic learning in innovative initiatives characterised by their experimental and co-creative nature, and which often face great uncertainty ([Arkesteijn et al, 2015](#)). DE is not about 'testing' a model of evaluation, but about generating it constantly. Therefore, this model is more appropriate than instrumental evaluation techniques to account for systemic change and deal with the unexpected and unpredictable.

Evaluation for strategic learning entails a process of acting, assessing, and acting again; it is an ongoing cycle of reflection and action. Strategic learning is a form of double loop learning. [Argyris and Schön \(1978\)](#) distinguish between single and double-loop learning. Single-loop learning takes place when an organisation detects a mistake, corrects it, and carries on with its present policies and objectives. Double-loop learning occurs when an organisation detects a mistake and changes its policies and objectives before it can take corrective actions. In strategic learning processes the integration of

explicit knowledge (codified, systematic, formal, and easy to communicate) and tacit knowledge (personal, context-specific and subjective) is crucial.

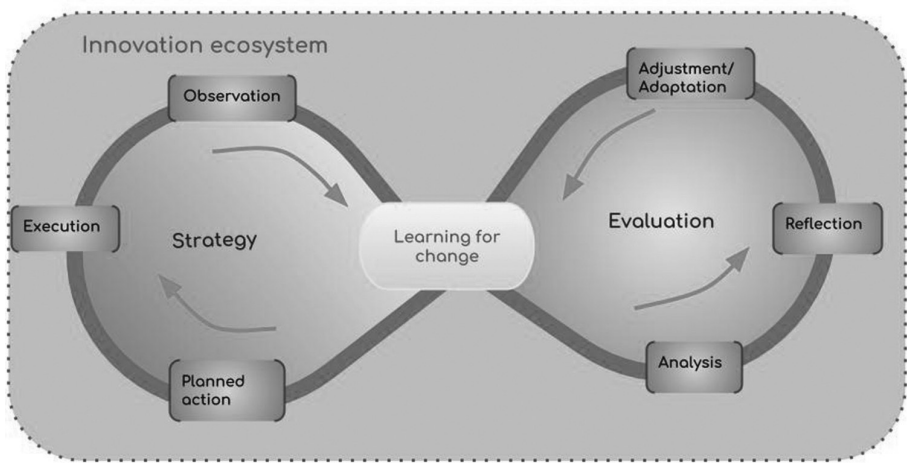
Based on this new evaluation paradigm, and the Lab’s concerns and requirements, we jointly decided that the evaluation proposal should contribute to its strategic learning process in order to improve systemic innovation. The Roadmap, validated not only with the Lab but also with AGESIC authorities, highlights the importance of processes of collaborative knowledge creation instead of focusing on objectives.

The rationale behind the pilot roadmap

The fundamental principle underlying this evaluation pilot is to assess the learning capacities for systemic change and to produce collective knowledge that supports the team when making decisions (Funtowicz and Ravetz, 1993; Gibbons et al, 1994). The evaluation seeks to address one of the organisation’s main challenges, such as the evaluation of co-creation processes and transdisciplinary knowledge generation, which is the basis of the Lab’s strategy (Bammer, 2005; Pohl et al, 2008; Lang et al, 2012; Polk, 2014). We therefore assumed that there is a reciprocal relationship between strategy and evaluation because, as observed by Preskill and Beer (2012, 4), when both elements are comprehended and carried out in this way, the organisation ‘is better prepared to learn, grow, adapt and continuously change in meaningful and effective ways’.

In this Roadmap, we assume there are two major interconnected phases in the learning process occurring within the innovation ecosystem where the Lab is inserted (Figure 1). The first phase involves the evolution of the strategy in which, after the execution of the planned actions, the team observes how the planned strategy unfolded. From the observation emerges the second phase in which the team analyses and reflects about the process to identify which elements of the planned strategy could be realised, which could not and which new ones emerged to adjust and adapt future actions. The learning for systemic change results from the interconnectedness of this two-fold process. This is a living model rather

Figure 1: The innovation ecosystem



Source: own elaboration

than a static one, and is thought to be an integral part of the Lab's core tasks as it informs and supports continuous innovation.

As innovation takes place under uncertainty, the actions carried out are experimental and demand constant reflection to understand what is happening in the process. The double-loop learning cycle is crucial to have the necessary information about how the strategy is unfolding and the resulting lessons (for example, to develop new tools for citizen engagement and participation). The evaluation is intentional about the use of data in a meaningful way to inform the process of innovation and identify emerging patterns and learn.

Thinking about the conditions and capacities for learning must be, therefore, the first and constant exercise of the Lab's team. Reflecting while the actions unfold (barriers and opportunities), enables generation of the necessary adjustments and adaptations to change the rules of the game. In this process, specific interventions can result in new rules, practices and relationships within the organisation and the network of actors involved (van Mierlo et al, 2010a; 2010b; 2013). Therefore, system learning needs to assess whether the current and relatively stable set of social structures is being challenged (Van Mierlo, 2010a), and what new knowledge, actions and practices are emerging.

The roadmap is 'co-adaptive' in order to meet the team wherever they may be in the project cycle and fit easily into the Lab's organisational process. The guidelines included in the prototype were developed to have a more purposeful and nimble impact to support projects with timely feedback. Consequently, the Lab can improve methods and tools by integrating emerging information in continuously changing environments.

The Roadmap design and validation process

A critical challenge when designing the Roadmap was to ensure its integration into the Lab's routines. To secure the appropriation of the new tool by the team, the design process underwent a series of stages: first, based on the review of literature and interviews with the team, we designed the pilot prototype presented in the previous section. Second, the pilot was analysed in three consecutive workshops in order to warrant the feedback process. In these events, apart from the Lab team, relevant stakeholders (AGESIC authorities, practitioners in the field of social and public innovation, and Lab workshops attendees) were invited. In total, 32 people participated in the three workshops (12, 10 and 10 respectively).

For the first workshop, held in October 2017, a document, containing the conceptual framework supporting the prototype, was created and discussed previously with the Lab team. During the workshop, the attendees worked on understanding the rationale behind the proposed tool, and on identifying the Lab's conditions and capacities for learning. Some of the triggers for the discussion were: What have we learnt from the co-creation processes and what outcomes can be connected to decision making? What evidence would indicate that a Lab's co-creation process is working or not? What have we learnt from our success and failures? What are the real-time feedback mechanisms of the organisation to track changes? What unforeseen events occurred and how did we respond to them?

A second workshop took place in March 2018, with the goal of reflecting on the system where the Roadmap would be immersed and the capacities and conditions of the Lab for learning. First, we focused on the innovation ecosystem by paying

close attention to current conditions in its subsystems: a) the institutional-political context; b) the innovative culture subsystem; c) the Lab-experimental strategy. For each of these subsystems, we proposed a series of trigger questions for their assessment. For the institutional-political context subsystem, the key issues to answer were: What institutional and systemic factors, such as policies, regulations, resource flows and administrative practices need to be in place to support, expand and sustain innovation? What are the obstacles and incentives for public innovation (in the AGESIC and with other sectors of government)? How is innovation organised to work collaboratively with other actors, government officials, citizens and the private sector? For the innovative culture subsystem, the key questions were: What cultural attributes, beliefs, narratives and values are required for public innovation to thrive (in the AGESIC and with other sectors of government)? Where do they exist and where do they meet resistance? For the Lab-experimental strategy subsystem, the key questions were: Do we have a clear strategy for the Lab? What aspects of the plan could be executed? What had to be ruled out and why? What adaptations have been made? What resources, skills, networks and knowledge are required in the public sector to support the scaling of innovation?

From this collective discussion, the Lab team proposed to introduce a qualitative self-evaluation online form to systematise the information of the ecosystem and to analyse it in their annual meetings with AGESIC authorities. The members of the team would answer the questionnaire, and the set of questions discussed during the workshop were reduced and simplified to the following (aiming to cover each subsystem): 1) How does the political and legal context hinder or encourage innovation in the Lab? 2) Have we developed a shared language and vision concerning the Lab practices? 3) Concerning the strategy the three critical questions are: 1) Why are we practising public innovation? 2) How are the short- and medium-term objectives linked to the theory of change? 3) What actions do we need to take to achieve these short- and medium-term objectives?

Finally, a third workshop took place in April 2018. After iteration and adaptation of the tool based on the feedback received on the first and second workshop, as well as on the continuous dialogue with the Lab team, we focused, on this occasion, on the validation of the prototype. This workshop was conducted using a design game method developed by the Lab to simulate the process of evaluation (Figure 2). Design games are a form of instrumental gaming data in experimental and innovative contexts for creating a common language. The activities usually involve relevant stakeholders and end users in both product and service design processes, through dialogue material to improve creativity (Vaajakallio, 2012). This board game was initially created with the aim of building co-creation capabilities among public servants. Through the game, the players would become familiar with the different stages of the co-creation process.

We adapted the board game and used it in the workshop to identify in which parts of such process the Roadmap should be integrated and how. As a result, it was recognised that it should play a critical role in monitoring the strategy while it is unfolding and assessing the experiments conducted by the Lab (for example, the introduction of new co-creation methods and tools). With regard to the experiments, the Roadmap serves to assess the capacities for learning by focusing on two outcomes: internal learning, aiming to capture the Lab team's reflections at the beginning and end of each co-creation project, and external learning seeking to apprehend the participants' reflections before and after attending the co-creation workshops.

Figure 2: Design Game



Source: the authors

For internal monitoring of learning capacities, an online protocol for each experiment will be implemented. It is to be responded to by members of the team involved in each project, and it contains the following information: 1) General project information (name of the experiment, participants, sector, time of development); 2) Pre-experiment questions: What is the activity that we are going to develop and why? What do we expect will happen (hypothesis)? How are we going to do it (method, tools)?; 3) Post-experiment questions: What unforeseen factors emerged and how did we adapt to them?

For the external evaluation of learning capacities, participants of the workshops will be required to answer an online questionnaire with the following questions: How useful was the activity? Did the co-creation process seem appropriate for the problem/s to be addressed? Was there an adequate treatment (respectful, humble, inclusive) of emerging ideas during the workshop by the Lab team? As an incentive for participants' response to the evaluation, they will receive an attendance certificate.

In synthesis, the process of designing the Roadmap led not only to the development of a monitoring and evaluation tool tailored to the organisation, but also to acknowledge the necessity for a more robust information system for decision making. The Roadmap as an information system per se will be complemented with online protocols and questionnaires to assess the Lab's innovation ecosystem, strategy and experiments.

Discussion and conclusions

This type of evaluation in the public sector is, without a doubt, a significant challenge since it introduces new rationales and forms of dialogue to those that bureaucrats and authorities are used to. However, it is precisely the role of an innovation Lab to take the lead in this process of change to legitimise experimental practice contexts in the public sphere.

In public experimentation, tension often arises from the demand of being accountable to authorities by measuring and communicating results. In this case, results are the product of learning processes that create value for the public sector. The Uruguayan Lab, before the development of the Roadmap, had only applied instrumental evaluation focusing on co-creation outputs. For example, instrumental evaluation was used in one of the most important projects of the Lab, 'Online Procedures', where co-creation was evaluated by the number of workshops, number of people who participated, number of prototypes and new online procedures created. For the members of the Lab, this form of evaluation contributed little to learning about the co-creation processes they had conducted, and about the new tools created or adapted. Therefore, by only using this model, the Lab was not evaluating the most critical outcome, which is the process that leads to changing the way of designing services and policies.

The Roadmap we proposed and validated with the Lab highlights the importance of the learning process instead of focusing on objectives. Although for a long time instrumental and learning evaluations were considered as antagonistic, recently diverse approaches ([Arkesteijn et al, 2015](#); [Regeer et al, 2009; 2016](#); [Taanman, 2014](#); [de Wildt-Liesveld et al, 2015](#); [Van Mierlo et al; 2010a; 2010b](#)) emphasise the need for their articulation from a systemic perspective. According to these approaches, the assessment of learning in complex systems involves reflecting and measuring non-linear processes of change with feedback loops and intertwined influence factors ([Cabrera et al, 2008](#); [Kurtz and Snowden, 2003](#); [Rogers, 2008](#); [Williams and Imam, 2007](#)).

The central contribution of our proposal was to reflect on the Lab's information system, by providing a tool paired with complementary instruments (online protocols and questionnaires) to acquire and use new data, provide feedback for the development of new ideas and tools, and support the capacity for innovation. At the same time, it is an easy and rapid method that will enable the Lab to monitor and evaluate the innovation ecosystem, where it is immersed, and its strategy, as well as to assess the outcomes from the experiments, improving accountability and communication. The Lab is not only a space 'for' experimentation but must think of itself as an experimental organisation. For this reason, we used the term 'Roadmap' instead of model because it is meant to be a heuristic tool rather than a set of prescriptive instructions about how to evaluate public innovation. It is a conceptual framework for learning and adaptation.

Acknowledgements

This article has been developed as part of the project 'Institutional Strengthening of the Social Innovation Lab for Digital Government', supported by the Inter-American Development Bank (No ATN/ AA-15240-UR). This contract is part of non-reimbursable technical cooperation between the IADB and the Uruguayan Government, aiming to increase the Lab's scope and impact and lay the foundations

for its long-term sustainability. The authors would like to thank the Lab team for their active participation and contribution to the documents created.

Conflict of interest

The authors declare that there is no conflict of interest.

References

- Agger, A., Sørensen, E. (2014) 'Designing collaborative policy innovation: lessons from a Danish municipality', in C. Ansell, J. Torfing (Eds.), *Public Innovation through Collaboration and Design*, London: Routledge, pp 188–208
- Agger, A. and Lund, D. H. (2017) 'Collaborative innovation in the public sector – new perspectives on the role of citizens?', *Scandinavian Journal of Public Administration*, 21(3), pp 17–37
- Ansell, C., Torfing, J. (2014) *Public Innovation through Collaboration and Design*, London: Routledge
- Ansell, C. K., Bartenberger, M. (2016) 'Varieties of experimentalism', *Ecological Economics*, 130, 64–73 doi: [10.1016/j.ecolecon.2016.05.016](https://doi.org/10.1016/j.ecolecon.2016.05.016)
- Antadze, N., Westley, F. R. (2012) 'Impact metrics for social innovation: Barriers or bridges to radical change?', *Journal of Social Entrepreneurship*, 3(2), 133–150 doi: [10.1080/19420676.2012.726005](https://doi.org/10.1080/19420676.2012.726005)
- Argyris, C., Schön, D. A. (1978) *Organizational Learning: A Theory of Action Perspective*, Reading, Mass: Addison-Wesley
- Argyris, C., Schön, D. A. (1996) *Organizational Learning II?: Theory, Method and Practice*, Reading, Mass: Addison-Wesley, Print
- Arkesteijn, M., van Mierlo, B., Leeuwis, C. (2015) 'The need for reflexive evaluation approaches in development cooperation', *Evaluation*, 21(1), 99–115 doi: [10.1177/1356389014564719](https://doi.org/10.1177/1356389014564719)
- Bammer, G. (2005) 'Integration and implementation sciences: Building a new specialization', *Ecology and Society*, 10, 2 doi: [10.5751/ES-01360-100206](https://doi.org/10.5751/ES-01360-100206)
- Bammer, G. (2017) 'Strengthening community operational research through exchange of tools and strategic alliances', *European Journal of Operational Research*, 0, 1–10
- Bason, C. (2010) *Leading Public Sector Innovation: Co-Creating for a Better Society*, Bristol: Policy Press
- Bason, C. (2014) *Design for Policy*, Farnham, Surrey: Routledge
- Bason, C. and Schneider, A. (2014) 'Public Design in Global Perspective; Empirical Trends' In Bason C. *Design for Policy*, Burlington, VT: Gower Pub Co.
- Bekkers, V., Edelenbos, J., Steijn, B. (2011) 'An innovative public sector? Embarking on the innovation journey', in V. Bekkers, J. Edelenbos, B. Steijn (Eds.), *Innovation in the Public Sector: Linking Capacity and Leadership*, New York: Palgrave Macmillan
- Blomqvist, K., Levy, J. (2006) 'Collaboration capability a focal concept in knowledge creation and collaborative innovation in networks', *International Journal of Management Concepts*, 2(1), 31–48
- Bund, E., Gerhard, U., Hoelscher, M., Mildemberger, G. (2015) 'A methodological framework for measuring social innovation', *Historical Social Research/Historische Sozialforschung*, 40(3):48–78
- Cabrera, D., Colosi, L., Lobdell, C. (2008) 'Systems thinking', *Evaluation and Program Planning*, 31(3), 299–310 doi: [10.1016/j.evalprogplan.2007.12.001](https://doi.org/10.1016/j.evalprogplan.2007.12.001)

- Dayson, C. (2016) 'Evaluating social innovations and their contribution to social value: The benefits of a 'blended value' approach', *Policy & Politics*, 45(3), 395–411
- De Wildt-Liesveld, R., Regeer, B.J. and Bunders, J. (2015) 'Governance strategies to enhance the adaptive capacity of niche experiments', *Environmental Innovation and Societal Transitions*, 16, 154–172 doi: [10.1016/j.eist.2015.04.001](https://doi.org/10.1016/j.eist.2015.04.001)
- Dewey, J. (1927) *The Public and Its Problems*, New York: H. Holt and Company
- Dozois, E., Blanchet-Cohen, N., Langlois, M. (2010) *A Practitioner's Guide to Developmental Evaluation*, International Institute for Child Rights and Development, University of Victoria, Canada
- Fuller, M., Lochard, A. (2016) 'Public Policy Labs in European Union Member States' (June): 22, <https://blogs.ec.europa.eu/eupolicylab/files/2016/10/Mapping-policy-labs-in-EU-MS.pdf>
- Funtowicz, S. O., Ravetz, J. R. (1993) 'Science for the post-normal age', *Futures*, 25(7), 739–755 doi: [10.1016/0016-3287\(93\)90022-L](https://doi.org/10.1016/0016-3287(93)90022-L)
- Gamble, J. A. (2008) *A Developmental Evaluation Primer*, Montreal: JW McConnell Family Foundation
- Gibbons, M., Limoges, C., Nowotny, H., Schwartzman, S., Scott, P., Trow, M. (1994) *The New Production of Knowledge: The Dynamics of Science and Research in Contemporary Societies*, London: Sage
- Graddy-Reed, A., Feldman, M. (2015) 'Stepping up: an empirical analysis of the role of social innovation in response to an economic recession', *Economy and Society*, 8(2), 293–312
- Hartley, J., Sørensen, E., Torfing, J. (2013) 'Collaborative innovation: A viable alternative to market competition and organizational entrepreneurship', *Public Administration Review*, 73, 821–830 doi: [10.1111/puar.12136](https://doi.org/10.1111/puar.12136)
- Hellstrom, T. (2013) 'Centre of excellence as a tool for capacity building', Draft Synthesis Report
- Hellstrom, T. (2015) 'Formative evaluation at a transdisciplinary research center', in M. Polk (Eds.), *Co-Producing Knowledge for Sustainable Cities: Joining Forces for Change*, London: Routledge, pp 146–165
- Junginger, S. (2017) 'Design research and practice for the public good: a reflection', *The Journal of Design, Economics, and Innovation* 3(4), 290–302 doi: [10.1016/j.sheji.2018.02.005](https://doi.org/10.1016/j.sheji.2018.02.005)
- Kimbell, L. (2016) 'Design in the Time of Policy Problems. Proceedings of DRS 2016, Design Research Society 50th Anniversary Conference. Brighton, UK, 27–30 June 2016
- Klein, J. T. (2008) 'Evaluation of interdisciplinary and transdisciplinary research', *American Journal of Preventive Medicine*, 35(2), 116–124. doi: [10.1016/j.amepre.2008.05.010](https://doi.org/10.1016/j.amepre.2008.05.010)
- Kurtz, C., Snowden, D. (2003) 'The new dynamics of strategy: sense-making in a complex and complicated world', *IBM Systems Journal*, 42(3), 462–483 doi: [10.1147/sj.423.0462](https://doi.org/10.1147/sj.423.0462)
- Lang, D. J., Wiek, A., Bergmann, M., Stauffacher, M., Martens, P., Moll, P., Thomas, C. J. (2012) 'Transdisciplinary research in sustainability science: Practice, principles, and challenges', *Sustainability science*, 7(1), 25–43 doi: [10.1007/s11625-011-0149-x](https://doi.org/10.1007/s11625-011-0149-x)
- McGann, M., Blomkamp E., Lewis, J. (2018) 'The rise of public sector innovation labs: Experiments in design thinking for policy', *Policy Sciences* 51(3), 249–267 doi: [10.1007/s11077-018-9315-7](https://doi.org/10.1007/s11077-018-9315-7)
- Midgley, G. (2003) *Systems Thinking*, London: Sage

- Milbergs, E., Vonortas, N. (2004) *Innovation metrics: Measurement to insight*, Center for Accelerating Innovation and George Washington University, National Innovation Initiative 21st Century, Working Group, 22
- Miyaguchi, T., Uitto J. I. (2017) 'What do evaluations tell us about climate change adaptation? Meta-analysis with a realist approach BT', in J.I. Uitto, J. Puri, R.D. van den Berg (Eds.), *Evaluating Climate Change Action for Sustainable Development*, Cham: Springer International Publishing, pp 235–254
- Morris, L. (2011) *The Innovation Master Plan: The CEO's Guide to Innovation*, Innovation Academy Walnut Creek, California.
- Mulgan, G. (2014) *The Radical's Dilemma: An Overview of the Practice and Prospects of Social and Public Labs*, https://media.nesta.org.uk/documents/social_and_public_labs_-_and_the_radicals_dilemma.pdf
- OECD (2017) *Fostering Innovation in the Public Sector*, Paris: OECD, <https://doi.org/10.1787/9789264270879-en>
- OECD (2018) *Embracing Innovation in Government: Global Trends 2018*, Paris: OECD
- Patton, M. Q. (2010) *Developmental Evaluation: Applying Complexity Concepts to Enhance Innovation and Use*, New York -London: Guilford Press
- Patton, M. Q. (2011) *Essentials of Utilization-Focused Evaluation*, Thousand Oaks, CA: Sage Publications
- Patton, C.V., Sawicki, D.S., Clark, J.J. (2013) *Basic Methods of Policy Analysis and Planning* (3a ed), Pearson
- Patton, M. Q. (2015) 'The sociological roots of utilization-focused evaluation', *The American Sociologist*, 46(4), 457–462. doi: [10.1007/s12108-015-9275-8](https://doi.org/10.1007/s12108-015-9275-8)
- Peters, G., Rava N. (2017) *Policy Design: From Technocracy to Complexity, and Beyond*, <http://www.ippapublicpolicy.org/file/paper/5932fa23369d0.pdf>
- Pohl, C., van Kerkhoff, L., Bammer, G., Hirsch Hadorn, G. (2008) 'Integration', G. Hirsch Hadorn, H. Hoffmann-Riem, S. Biber-Klemm, W. Grossenbacher Mansuy, D. Joye, C. Pohl, U. Wiesmann, E. Zemp (Eds.) *Handbook of Transdisciplinary Research*, Dordrecht: Springer, pp 411–424
- Polk, M. (2014) 'Achieving the promise of transdisciplinarity: A critical exploration of the relationship between transdisciplinary research and societal problem solving', *Sustainability Science*, 9(4), 439–451 30. doi: [10.1007/s11625-014-0247-7](https://doi.org/10.1007/s11625-014-0247-7)
- Preskill, H., Beer, T. (2012) *Evaluating Social Innovation*, https://guelph.ca/wp-content/uploads/Contract_12-191_Description.pdf
- Puttick, R., Baeck, P., Colligan, P. (2014) 'i-teams: the teams and funds making innovation happen in governments around the world', Nesta and Bloomberg Philanthropies, www.theiteams.org
- Regeer, B.J., de Wildt-Liesveld, R., van Mierlo, B., Bunders, J. F. (2016) 'Exploring ways to reconcile accountability and learning in the evaluation of niche experiments', *Evaluation*, 22(1), 6–28 doi: [10.1177/1356389015623659](https://doi.org/10.1177/1356389015623659)
- Regeer, B.J., Hoes, A.-C., Amstel-van Saane, M. van, Caron-Flinterman, F., Bunders, J. (2009) 'Six guiding principles for evaluating mode-2 strategies for sustainable development', *American Journal of Evaluation*, 30(4), 515–537 doi: [10.1177/1098214009344618](https://doi.org/10.1177/1098214009344618)
- Rogers, P.J. (2008) 'Using programme theory to evaluate complicated and complex aspects of interventions', *Evaluation* 14(1), 29–48 doi: [10.1177/1356389007084674](https://doi.org/10.1177/1356389007084674)
- Scriven, M. (1996) 'Types of evaluation and types of evaluator', *Evaluation Practice*, 17(2), 151 doi: [10.1016/S0886-1633\(96\)90020-3](https://doi.org/10.1016/S0886-1633(96)90020-3)

- Snowden, D.J., Boone, M. E. (2007) A leader's framework for decision making, *Harvard Business Review*, 85(11), 68–77
- Stringer, E. T. (2007) *Action Research*, Los Angeles: Sage Publications
- Taanman, M. (2014) *Looking for Transitions*, Erasmus University Rotterdam
- Timeus, K., Gascó, M. (2018) 'Increasing innovation capacity in city governments: Do innovation labs make a difference?', *Journal of Urban Affairs*, (March), 40(7), 992–1008
- Tönurist, P., Kattel, R., Lember, V. (2017) 'Innovation labs in the public sector: What they are and what they do?', *Public Management Review*, 19(10), 1455–1479 doi: [10.1080/14719037.2017.1287939](https://doi.org/10.1080/14719037.2017.1287939)
- Vaajakallio, K., Mattelmäki, T. (2014) 'Design games in codesign: as a tool, a mindset and a structure', *CoDesign*, 10(1), 63–77 doi: [10.1080/15710882.2014.881886](https://doi.org/10.1080/15710882.2014.881886)
- Van de Poel, I., Asveld, L., Mehos, D. C. (Eds.). (2017) *New Perspectives on Technology in Society: Experimentation Beyond the Laboratory*, Routledge
- Van Mierlo, B. C., Arkesteijn, M., Leeuwis, C. (2010a) 'Enhancing the reflexivity of system innovation projects with system analyses', *American Journal of Evaluation*, 31(2), 143–161 doi: [10.1177/1098214010366046](https://doi.org/10.1177/1098214010366046)
- Van Mierlo, B. C., Janssen, A. P., Leenstra, F. R., Van Weeghel, H. J. E. (2013) 'Encouraging system learning in two poultry subsectors', *Agricultural Systems*, 115, 29–40 doi: [10.1016/j.agsy.2012.10.002](https://doi.org/10.1016/j.agsy.2012.10.002)
- Van Mierlo, B. C., Regeer, B., van Amstel, M., Arkesteijn, M. C. M., Beekman, V., Bunders, J. F. G., de Cock Buning, T., Hoes, A. C., Leeuwis, C. (2010b) 'Reflexive monitoring in action. A guide for monitoring system innovation projects', *Communication and Innovation Studies*, WUR, Wageningen/Amsterdam: Athena Institute
- Voorberg, W. H., Bekkers, V. J., Tummers, L. G. (2015) 'A systematic review of co-creation and co-production: Embarking on the social innovation journey', *Public Management Review*, 17(9), 1333–1357 doi: [10.1080/14719037.2014.930505](https://doi.org/10.1080/14719037.2014.930505)
- Westley, F., Zimmerman, B., Patton, M. (2006) *Getting to Maybe: How the World is Changed*, Canada: Vintage
- Williams, B., Imam, I. (2007) *Systems Concepts in Evaluation: An Expert Anthology*, Point Reyes, CA: Edge Press/American Evaluation Association