

# The hidden side of an evaluation culture:

The importance of assessing employee's support to core evaluation values in public organizations

Daniel Grimaldi\*  
[dgrimald@gmu.edu](mailto:dgrimald@gmu.edu)

## Abstract

The use of rigorous and systematic methods to evaluate policies is still timid. To understand why it is crucial to understand the role of evaluation cultures inside public organizations. On a first and visible tier of the evaluation culture lays the set of norms, rules, and technical agreements that try to enforce parameters for the execution and dissemination of monitoring and evaluation (M&E) activities. This first component is usually referred to as the organization's M&E system. On a deeper level, there are values and beliefs that guide the employees' attitudes toward the M&E system. This work defines an M&E system and characterizes how two public organizations (IDB and BNDES) have implemented it. Still, to fully understand an evaluation culture it is necessary to assess how critical values of the organization interact and support its M&E system. To tackle this challenge, this work proposes a three-steps approach for a full assessment of evaluation cultures in public organizations. Based on structured interviews, desk research, and simulation exercises, the proposed method should be able not only to characterize each organization's M&E system but also to identify the acceptance of its employees to three core evaluation values.

## Introduction

“(...) some evaluation advocates appear to believe that if high-quality, rigorous evaluations are conducted their findings and recommendations will almost automatically be used because of their self-evident value. Anyone who has worked at a senior level in government will recognize the fallacy in this belief.” (Lopez-Acevedo et al., 2012, p. 6)

The importance attributed to monitoring and evaluation (M&E) activities in modern western countries has been increasing steadily, at least since the 1980s (Leeuw and Furubo, 2008).

From a technical perspective, this movement has two key supporters. The first is the improvement in computational capacity. The second is the advances in statistics, giving rise

---

\* PhD Student at the Schar School of Policy and Government of the George Mason University.

to a new causal science (Pearl et al., 2016). In private organizations, the use of large datasets and evaluation techniques has been embraced as a powerful tool to increase sales and profits. Recently, an influential business magazine has even decreed data as the new oil (The Economist, 2017).

Despite that, the use of rigorous and systematic methods to evaluate policies is still timid. To understand why it is useful to consider that the barriers to adopting M&E activities inside organizations can have two distinct natures. On the supply side, one must have access to datasets, computers, and technical skills. On the demand side, the organizational culture must hold norms and processes that encourage its employees to use these inputs, as well as values and beliefs that embrace the appropriation of evaluation findings.

As pointed out in the epigraph of this section, in general, the binding constraints for the use of rigorous evaluations' practices are not on the supply side anymore. To understand the chains that have been holding the advances of M&E activities in the public sector, it is crucial to understand the role of evaluation cultures inside public organizations.

Following Forss (2012) and Mayne (2012), evaluation culture will be defined here as a part of the organizational culture. As such, it has two broad components. On a first and visible tier, lays the set of norms, rules and technical agreements that try to enforce parameters for the execution and dissemination of M&E activities inside an organization. This first component is usually referred to as M&E systems (Leeuw and Furubo, 2008). On a deeper level, there are values and beliefs that guide the employees' attitudes toward the M&E system.

Several authors have studied evaluation cultures (Barbier and Hawkins, 2012; Furubo et al., 2002; Lopez-Acevedo et al., 2012). This literature, however, has focused on its first layer.

Inspired by Baekgaard et al.(2019) and Banuri et al.(2019), this research proposes a new methodology to capture not only the observable characteristics of different M&E systems, but also the hidden elements of an evaluation culture.

To do so, this proposal is divided into four sections, besides this introduction. The next one will define M&E systems and discuss two particular cases of public organizations that have implemented them: the Inter-American Development Bank (IDB) and the Brazilian Development Bank (BNDES). The third section will defend why the mere observation of formal norms and processes related to M&E activities is not enough to fully comprehend an evaluation culture. While important, institutionalization is not all that matters. Then the fourth section will offer three crucial values that should be present in organizational culture to support M&E activities. Finally, the last part brings a proposal for a full assessment of evaluation cultures in public organizations, as well as an idea for a first application with IDB, BNDES, and a representative commercial bank.

## **Institutionalization matters**

The implementation of M&E activities on a larger scale requires the establishment of norms and rules to coordinate tasks that were previously done through *ad hoc* initiatives. When this coordination effort is institutionalized, it gives rise to evaluation systems.

According to Leeuw and Furubo (2008) M&E systems are a collections of norms, rules, techniques and actors arranged under two specific mandates: i) to define, calculate, and publicize inputs and output measures, and ii) to rigorously and systematically evaluate the underline causes of an observed performance – this causal assessment will be here referred to as impact evaluation.

This second aim is particularly relevant to public organizations because the connection between policy efforts and their impacts on society's well-being is not easily identifiable. However, while performance monitoring can usually be done by typical administrative staff, execution of impact evaluations requires a more specific set of skills, mixing semi-academic and semi-administrative capabilities. In part because of that, though not solely, M&E systems are less frequently observed than SPMs (Gaarder and Briceño, 2010). The IDB and the BNDES are two examples of public organizations that have adopted them.

### **The IDB M&E system<sup>1</sup>**

The Inter-American Development Bank (IDB) has a worldwide recognized M&E system. IDB is a multilateral organization whose main declared objectives are: i) reduction of poverty and social inequalities; ii) enhancement of productivity growth and innovation; and promotion of regional economic integration of vulnerable countries. The complexity of assessing its contributions to such aims, as well as external pressures from donator countries, may explain why IDB has developed complex evaluation norms.

At the top of IDB's M&E system is the Board of Governors. The member countries nominate its participants – normally, among national treasury ministers or presidents of central banks. This governing body is responsible for overseeing the work of the Board of Executive Directors, who conduct daily decisions and approve banks' operations. The IDB's president is elected by the Board of Governors and runs the directors.

Every IDB operation passes through several layers of analysis. Beyond traditional financial and risk assessments, projects are also examined through a lens of effectiveness – which

---

<sup>1</sup> This section is based on an unstructured interview conducted with an employee of the Office of Evaluation and Oversight (OVE), as well as on the website of IDB ([www.idb.org](http://www.idb.org)).

means their contribution to IDB's development goals are also evaluated. In order to do so, IDB has institutionalized an assessment tool called Development Effectiveness Matrix (DEM)<sup>2</sup>.

IDB also has the Strategic Planning and Development Effectiveness Division (SPD), a unit dedicated to effectiveness evaluation. As such, SPD is also responsible for helping operational teams applying the DEM, and for building tools and norms that the operational team can use across the project cycle.

The project cycle starts with the analysis, but also includes disbursement, monitoring, and amortization phases. IDB has built norms that oblige his employees to monitor and evaluate not only the repayment of loans but also the effectiveness of funded projects<sup>3</sup>. Operational teams conduct all this process with the help of SPD.

Besides helping operational teams, SPD also conducts impact evaluations for selected operations. SPD, however, answers to the Board of Directors, like any other operational unit. As such, it might suffer pressures to bias its effectiveness analysis in order to meet the financial objectives of the institution. IDB is still a bank, and it must fulfill financial targets that could antagonize development goals. After all, the more profitable projects might not be the most effective ones.

It is worth noting that SPD might resist any attempt to influence the outcomes of evaluations. Still, the bare possibility of undue interference is enough to undermine the credibility of

---

<sup>2</sup> DEM is a checklist, where grades are attributed to the following dimensions: (i) strategic alignment with IDB's development goals; (ii) strategic alignment with the hosting country goals; (iii) theory of change; (iv) M&E design; (v) economic return; (vi) risks; and (vii) additionality.

<sup>3</sup> Besides DEM, there are several different reports operational teams must comply with during the life cycle of a projects, including the Project Monitoring Report (PMR) and the Loan Results Report (LRR). Both have elements of financial and effectiveness assessments.

results found by SPD. Peer reviews and validation from experts can also compensate for this. Nonetheless, the general suspicion with results found by SPD's activities will tend to be higher due to these conflicting interests.

It helps explain why the formal independence of the evaluation units is a key feature for every M&E system.<sup>4</sup> When the main objectives are internal learning and managerial decisions, decentralized units attached to operational areas are a more suitable approach for the M&E system. On the contrary, when evaluation activities are more worried about providing transparency to external actors and information for strategic budgetary decisions, there will be a higher pressure for formal independence.

To cope with such pressures, IDB established in 1999 the Office of Evaluation and Oversight (OVE). OVE is an independent unit responsible for conducting an impact evaluation agenda and for reviewing the work of operational teams and SPD through meta-analysis. The Director of OVE is an external professional, hired for five-years fixed terms, and he answers directly to the Board of Governors.

### **The BNDES M&E system<sup>5</sup>**

The Brazilian Development Bank (BNDES) has been operating with long-term credit lines since 1952. Its actual purpose is to “transform the life of Brazilian generations, promoting sustainable development”<sup>6</sup>. Again, to assess an institution's contribution to such an aim requires proper evaluation skills. Still, BNDES' M&E system is very recent. Only in 2015

---

<sup>4</sup> According to OECD (2002), an evaluation is independent when it is executed by “entities and persons free of the control of those responsible for the design and implementation of the development intervention.” (p.24)

<sup>5</sup> This section is based on an unstructured interview with an employee of the Department of Evaluation and Effectiveness Promotion (DEAPE), as well as on the website of BNDES ([www.bndes.gov.br/effectiveness](http://www.bndes.gov.br/effectiveness))

<sup>6</sup> <https://bit.ly/2OP5zRf>

BNDES approved a formal M&E policy. Norms establishing major evaluation practices for operational areas were implemented only in 2018.

The M&E systems of BNDES and IDB share plenty of similarities. In terms of decision structures, BNDES has as at the top the Board of Administration, which is responsible for the strategic guidance of the organization. Its members are all appointed by the Brazilian federal government and came from outside the BNDES' bureaucracy. There is also a Board of Executive Directors in charge of daily decisions, including approval of all BNDES' operations.

Since 2018, BNDES has also established specific procedures for assessing the effectiveness of the projects it funds. For a superficial ex-ante evaluation of effectiveness, it uses a tool called Thesis for the Impact on Projects Investments (Tiip). For registering expected development goals attached to each project, it uses the Results Frame (QR).

The Department of Evaluation and Effectiveness Promotion (DEAPE) was created in 2016 and is a specialized unit in charge of evaluation activities. It now counts with approximately 20 professionals and is responsible for helping operational areas with the following tasks: i) to assess the effectiveness of projects, during the analysis phase; ii) to monitor development outputs and outcomes during and after the conclusion of the investments; and iii) to implement ex-post impact evaluations for selected projects. Since 2018, after pressures from the Brazilian Accounting Court (TCU), all evaluation findings reached by DEAPE or by external evaluators hired by it must be published on the website of the institution.

The comparisons between DEAPE, SPD, and OVE stress the main differences between BNDES and IDB M&E systems. In BNDES, DEAPE plays two different roles. It manages the M&E system and implements impact evaluations, just as OVE does for the IDB.

However, it is as close to operational areas as is SPD, and, therefore, suffers from the same conflicting interests. As Chart 1 shows, the absence of an independent evaluation unit, reporting directly to the board in charge of strategic decisions, is the main difference between these two studied cases.

**Chart 1: Main features of BNDES and IDB M&E systems**

<b>Features</b>	<b>BNDES</b>	<b>IDB</b>
Dedicated unit inside the executive structure	Department of Evaluation and Effectiveness Promotion (DEAPE) is responsible for managing the M&E system, for implementing impact evaluations and for helping operational areas with their effectiveness assessments.	Strategic Planning and Development Effectiveness Division (SPD) is responsible for implementing impact evaluations, and for helping operational areas with their effectiveness assessments
Dedicated unit outside executive structure (independent unit).	There is no independent unit	Office of Evaluation and Oversight (OVE) is responsible for managing the M&E system and for implementing impact evaluations.
Operational teams' engagement with evaluation activities	Operational teams engage in evaluation assessments during the analysis and execution of the projects.	Operational teams engage in evaluation assessments during the analysis and execution of the projects
Evaluation activities executed within the M&E system	Output and outcomes monitoring, impact evaluations, and meta-analysis.	Output and outcomes monitoring, impact evaluations, and meta-analysis.
Communication strategy	Internal review process inside the executive structure, followed by publication on the website and on an annual report.	Internal review process, with evaluation findings being reported for the Board of Governors, followed by publication on the website and on an annual report.

Source: Author's elaboration



## **Evaluation culture: is institutionalization all that matters?**

The two cases presented here are not isolated experiences. The last two decades have witnessed the consolidation of M&E systems. By studying a sample of 19 developing countries, Jacob et al. (2015) concluded that 15 presented a high degree of maturity on its national M&E system, and the other 4 were on a medium level. Applying the same methodology roughly 10 years before, Furubo et al. (2002) found that only 10 had a mature M&E system, and 4 nations were still working on a low level of consolidation.

To reach that conclusion, these authors have used 9 different indicators, that can easily be aggregated into two broader categories: measures for the existence of norms and rules regarding evaluation procedures within governments<sup>7</sup> and measures for the supply of adequate technical skills for evaluation practices<sup>8</sup>.

By applying these same rules for the cases of BNDES and IDB, we would probably conclude that both institutions have M&E systems with similar maturity levels – especially because the criteria do not explicitly take into consideration the existence of an independent M&E unit. Still, evaluation activities have a much longer history in IDB than BNDES.

Following DiMaggio (1994), norms, rules, and techniques can be broadly understood as institutions, an essential component of a culture. Therefore, it is reasonable to interpret the consolidation of M&E systems as sign of the advancement of evaluation culture across

---

<sup>7</sup> The indicators used by Jacob et al. (2015, p. 8) that fit this category are: “institutional arrangements in the government for conducting evaluations and disseminating their results exist”, “institutional arrangements in Parliament for conducting and disseminating evaluations exists”; “evaluation activities occur within the supreme audit institution”, “evaluation takes place in many policy domains”, “discussions and debates fuel a national discourse regarding evaluation”,

<sup>8</sup> The indicators used by Jacob et al. (2015, p. 8) that fit in this second category are: “there should be a supply of evaluators specializing in different disciplines”, “national evaluation society exists”, “pluralism exists within each policy domain”, “evaluations do not just focus on inputs/outputs, but also on outcomes”.

organizations within the public sector of these nations. Nonetheless, institutionalization is not enough to operationalize the full concept of evaluation culture, which must also incorporate values and beliefs (Forss, 2012; and Mayne, 2012).

For some authors, institutionalization is the dominant aspect because if norms and rules are enforced for long enough, they will end up promoting inside the organization the required values and beliefs (Gaarder and Briceño, 2010). That view is also present on numerous case-studies that reviewed the consolidation M&E systems – see, for instance, (Furubo et al., 2002; U.S. GAO, 2003). In this literature, one can frequently find the suggestion of a causal relation, going from external pressures to the top-down enactment of institutional changes that, ultimately, lead to the spread of an evaluation culture.<sup>9</sup> It also seems to be the cases of IDB and BNDES.

Still, this is not the only available path. Norms and rules will interact with employees and external stakeholders' prior beliefs, and there can be resistance to the will of managers to impose certain practices. As a part of an organizational culture, the evaluation culture will be the result of daily interactions among all practitioners (Seel, 2005). In the long-run, this process can reinforce or weaken top-down established norms and rules (Mayne, 2009). That is why organizations with different sets of values, even when submitted to similar M&E systems, will not, necessarily, converge to produce the same evaluation culture.

---

<sup>9</sup> That is the dominant view presented by Furubo et al. (2002) when they studied national M&E systems. Also, U.S. GAO (2003) defend that external conditions, mostly policy debates and budget constraints, were major forces pushing the adoption of systematic evaluation activities in federal agencies of the United States.

## Defining core evaluation values

The notion that a few fundamental beliefs are responsible for persistent attitudes has been explored to explain political (Converse, 2006; Heath et al., 1994) and organizational (Cooke and Szumal, 1993) patterns. It is only intuitive that the same principle applies to evaluation activities. Still, what could be these core values, essential to favor M&E practices?

Maybe because of its strong emphasis on institutions, recent literature does not have a precise answer to that question. Nonetheless, it is possible to identify, across the work of different authors, some values that are typically observed in organizations that are recognized by having successful M&E systems.

First, these organizations **see merit on activities that support evaluation, such as quality control, data generation, and outspoken debate** (Barbier and Hawkins, 2012; Lopez-Acevedo et al., 2012; Mayne, 2009). That intrinsic value makes it easier for the employees to accept the institutionalization of evaluation practices, as well as to recognize the need to direct resources for that purpose.

Second, organizations with well-established evaluation cultures will show **commitment to self-examination or self-reflection** (Lopez-Acevedo et al., 2012; Mayne, 2009; U.S. GAO, 2003). This value will make employees more willing to question their *prior* beliefs and, therefore, to engage in evidence-based learning.

The third value commonly associated with evaluation cultures is the **commitment with permanent improvement** (Lopez-Acevedo et al., 2012; Mayne, 2009; United States U.S. GAO, 2003). M&E activities will likely bring evidence against organizational practices and routines. An understandable reaction would be to discard or discredit this evidence to

preserve the *status quo*. When employees are involved in achieving better policies, there will be a push against this conservative reaction, and evaluation findings will have a better chance to promote changes.

Consolidating this last core value seems to be particularly challenging for organizations in charge of policies. First, because it is hard to establish a clear contribution of a policy to distant outcomes – as we have seen for the IDB and BNDES cases. This makes it easier for political interests to resist and discredit evaluation findings.

Second, because outcome achievement is frequently disassociated with pecuniary gains for the employees. While the profit-signal can always force private companies to correct their practices, this kind of alignment mechanism is not usually available for public organizations. For several different reasons, public bureaucracy might try to resist policy changes even when they seem to be a path for more effective state action.

Despite their reasons, when employees of an organization are not committed to these three core values, they will probably resist norms trying to impose M&E systems. They can also comply with the rules loosely without never fully engaging in an evaluation agenda. Either way, the consolidation of an evaluation culture will only be achieved through a change in these organizational values.

Finally, other values are also mentioned in the literature, but with lower consistency. Among these, one deserves a note: the **commitment to independency for evaluation activities**. For some authors, this value is crucial to promote the use of evaluation findings (Lopez-Acevedo et al., 2012; U.S. GAO, 2003). However, for others, this is a mere external constraint, useful only when evaluation culture is weak (Mayne, 2012).

In organizations already committed to self-examination and improvement, formal independence would serve only to create an artificial distance between evaluators and policy practitioners, hurting the natural flow of tacit knowledge that could benefit both. On the other hand, when these core values are not present, formal independence can force employees to respect and reflect upon evaluation findings. In that sense, independency would be only a tool to promote evaluation values.

### **A full assessment of an evaluation culture**

This work intends to propose a three-steps approach for a full assessment of evaluation cultures in public organizations. The first step is based on a structured interview with closed and open-ended questions. It must be conducted with only a few high-rank employees of the organization under analysis. The outcome of this interview will guide a second phase that consists of desk research. The aim is to study norms and other official documents that characterize the organization's M&E system. Finally, the last step of the assessment will require the application of a structured simulation to a representative sample of employees. This exercise will serve to identify the acceptance of the employees to the core evaluation values.

Appendix A presents the 25-questions that will guide the structured interview phase. They are divided into four subsections designed to: i) understand the academic background and the professional duties of the interviewed; ii) identify norms and other formal documents that organization uses to direct evaluation activities; iii) explore which kind of organizational structure has evolved to deal with evaluation activities; and iv) recognize some core features

of the evaluation system, including goals, methods, consumers and the degree of operational engagement.

The norms and other formal documents identified during the interview phase will later be studied to validate initial perceptions. Following the experiences of Furubo et al. (2002) and Jacob et al. (2015), these initial steps are adequate to identify and characterize the first layer of an evaluation culture – that is, its M&E system.

The initial exploration would also serve to guide the third step of the methodology. Inspired by the works of Baekgaard et al.(2019) and Banuri et al. (2019), this last phase consists of two customized simulations. They were designed to grasp the hidden parts of the evaluation culture through the measurement of its employees' support to core evaluation values. These simulations would be applied through online survey tools to a representative sample of medium and low-rank staff inside the organization.<sup>10</sup>

Appendix B presents the three questions associated with these two simulations. The first one will evaluate how much the employees would be willing to invest in evaluation activities, with which purpose and how their response would react to different public contexts – more supportive or more critic to the policy being executed. It will serve to measure the presence of the first core evaluation value. The critical assumptions are that in organizations that see intrinsic merit on evaluation activities will be able to observe: i) a greater willingness to direct resources for evaluation efforts, and ii) inferior responsiveness to public contexts.

In the second simulation, employees will be presented to a policy that has clear development goals and strong public support, but no relevant evidence of its results to society. Important

---

<sup>10</sup> Examples of online survey tools that could be used are LimeSurvey (<https://www.limesurvey.org/>) or SurveyMonkey (<https://www.surveymonkey.com/>).

to say that half of the respondents will be randomly assigned to a policy that has no relation to the core business of his organization and will be required to picture himself as an ordinary citizen. The other half will be required to imagine themselves as managers of a policy that is close-related to the core-business of his organization.

After that, they will be informed about the existence of a causal evaluation done through a Randomized Control Trial (RCT). The desired outcome variable will be observed for control and treatment groups, and the values will clearly support a slightly negative impact. Within that context, employees will be required to answer two questions. The first consists of a mere interpretation of the results, while the second will test the support of the respondent to different strategies to deal with the evaluation findings.

The crucial assumptions behind the first question of this second simulation are that workers of an organization more committed to self-examination and evidence-based learning will: i) present a more accurate reading of the results; and ii) be less sensitive to confirmation bias (Banuri et al., 2019).

For the last question of the simulation, the hypothesis is that when an organization is more committed to self-improvement, its employees will, on average, deal with undesired results in a more positive way. In particular, they will be more supportive of strategies that favor results' transparency and policy change.

In principle, this methodology could be applied to a broad spectrum of public organizations. Still, as a way of calibrating the questionnaires and validating the overall structure, the proposal here is to perform a trial with undergraduate students, followed by a first round of application with three institutions: IDB, BNDES, and a private commercial bank.

IDB is widely recognized as an organization with a mature M&E system. BNDES is also a development bank, but it has only recently institutionalized its evaluation practices. Since commercial banks typically operate only with audit and performance monitoring systems (Leeuw and Furubo, 2008), the private commercial bank would serve as a benchmark without an M&E system.

This selection of participants is ideal for the validation of the methodology. Firstly, due to its development goals, one should expect that IDB and BNDES have more consolidated M&E systems when compared to the commercial bank. These differences should have clear impacts in terms of norms and organizational structure.

Besides that, private organizations are profit-driven and should not attribute intrinsic value to evaluation activities that cannot contribute to that objective. On the other hand, organizations with development goals, as IDB and BNDES, should recognize the value of M&E even when it implies restrictions to their profits.

Therefore, under the proposed simulation, one should expect that the commercial bank will have weak support for core evaluation values every time they conflict with profit maximization. Considering the two development banks, one should expect to observe higher support for core evaluation values among IDB employees because BNDES has only recently institutionalized its evaluation activities.

## References

- Baekgaard, M., Christensen, J., Dahlmann, C.M., Mathiasen, A., Petersen, N.B.G., 2019. The Role of Evidence in Politics: Motivated Reasoning and Persuasion among Politicians. *British Journal of Political Science* 49, 1117–1140. <https://doi.org/10.1017/S0007123417000084>
- Banuri, S., Dercon, S., Gauri, V., 2019. Biased Policy Professionals. *The World Bank Economic Review* 33, 310–327. <https://doi.org/10.1093/wber/lhy033>



- Barbier, J.-Claude., Hawkins, Penny. (Eds.), 2012. *Evaluation cultures: sense-making in complex times*, Comparative policy evaluation; v. 19. Transaction Publishers, New Brunswick.
- Converse, P.E., 2006. The nature of belief systems in mass publics (1964). *Critical Review* 18, 1–74. <https://doi.org/10.1080/08913810608443650>
- Cooke, R.A., Szumal, J.L., 1993. Measuring Normative Beliefs and Shared Behavioral Expectations in Organizations: The Reliability and Validity of the Organizational Culture Inventory. *Psychological Reports* 72, 1299–1330. <https://doi.org/10.2466/pr0.1993.72.3c.1299>
- DiMaggio, P., 1994. “Culture and Economy, in: Smelser, N.J., Swedberg, R. (Eds.), *The Handbook of Economic Sociology*. Princeton University Press, Princeton, NJ, pp. 27–57.
- Forss, K., 2012. Four organizations - four evaluation cultures, in: Barbier, J.-Claude., Hawkins, Penny. (Eds.), *Evaluation Cultures: Sense-Making in Complex Times*, Comparative Policy Evaluation; v. 19. Transaction Publishers, New Brunswick.
- Furubo, J.-Eric., Rist, R.C., Sandahl, Rolf. (Eds.), 2002. *International atlas of evaluation*, Comparative policy analysis series. Transaction Publishers, New Brunswick, U.S.A.
- Gaarder, M.M., Briceño, B., 2010. Institutionalisation of government evaluation: balancing trade-offs. *Journal of Development Effectiveness* 2, 289–309. <https://doi.org/10.1080/19439342.2010.505027>
- Heath, A., Evans, G., Martin, J., 1994. The Measurement of Core Beliefs and Values: The Development of Balanced Socialist/Laissez Faire and Libertarian/Authoritarian Scales. *British Journal of Political Science* 24, 115–132.
- Jacob, S., Speer, S., Furubo, J.-E., 2015. The institutionalization of evaluation matters: Updating the International Atlas of Evaluation 10 years later. *Evaluation* 21, 6–31. <https://doi.org/10.1177/1356389014564248>
- Leeuw, F.L., Furubo, J.-E., 2008. Evaluation Systems: What Are They and Why Study Them? *Evaluation* 14, 157–169. <https://doi.org/10.1177/1356389007087537>
- Lopez-Acevedo, G., Krause, P., Mackay, K., 2012. *Building better policies: the nuts and bolts of monitoring and evaluation systems* (No. 68166). The World Bank.
- Mayne, J., 2012. Independence in evaluation, in: Barbier, J.-Claude., Hawkins, Penny. (Eds.), *Evaluation Cultures: Sense-Making in Complex Times*, Comparative Policy Evaluation; v. 19. Transaction Publishers, New Brunswick.
- Mayne, J., 2009. Building an Evaluative Culture: The Key to Effective Evaluation and Results Management. *The Canadian Journal of Program Evaluation*; Toronto 24.
- OECD, O. for E.C. and D., 2002. Glossary on key terms in evaluation and results managed.
- Pearl, J., Glymour, M., Jewell, N., 2016. *Causal inference in statistics: a primer*, 1st ed. John Wiley & Sons, West Sussex, United Kingdom.
- Seel, R., 2005. Culture & Complexity: new insights on organizational change., in: *Organizational Culture: An Introduction*. ICFAI University Press.
- The Economist, 2017. The world’s most valuable resource is no longer oil, but data. *The Economist*.
- United States Government Accountability Office (U.S. GAO), 2003. An evaluation culture and collaborative partnership help build agency capacity (No. 454).

## Appendix A: Description of the structured interview

This questionnaire should be applied to high-rank employees responsible for managing monitoring and evaluation activities and to high-rank employees in charge of the execution of programs, policies, or other operational tasks that represent the core business of the organization.

<b>Section One - General profile of the interviewed employee:</b>
<b>1. Date of the interview (YYYY-MM-DD):</b>
<b>2. Full Name:</b>
<b>3. Number of full years working in the organization:</b>
<b>4. Current job title:</b>
<b>5. Name of the division / unit / department of the employee:</b>
<b>6. The core business of this division is closely related with (select all that apply):</b> <input type="checkbox"/> Execution of operational tasks related to the organization's core business <input type="checkbox"/> Management of operational tasks related to the organization's core business <input type="checkbox"/> Execution of performance monitoring activities <input type="checkbox"/> Execution of evaluation activities <input type="checkbox"/> Management of monitoring and evaluation activities <input type="checkbox"/> Execution of audit and compliance activities <input type="checkbox"/> Management of audit and compliance activities
<b>7. Educational background of the employee</b>  <b>7a. Highest academic title:</b> <input type="checkbox"/> High school <input type="checkbox"/> Undergraduate <input type="checkbox"/> Master's degree <input type="checkbox"/> Ph.D. or other Doctoral degree  <b>7b. Specialization field of the highest academic title:</b> <input type="checkbox"/> Economics <input type="checkbox"/> Political Science <input type="checkbox"/> Statistics <input type="checkbox"/> Public policy <input type="checkbox"/> others - specify: _____

<p><b>Section Two – Overview of norms related to evaluation activities</b></p>
<p><b>8. Does your institution possess a formal policy establishing principles for evaluation activities?</b></p> <p>( ) Yes ( ) No ( ) I do not know</p> <p><i>If the answer is “No” or “I do not know”, jump to question 10</i></p>
<p><b>9. Is the policy mentioned on question 8 a public document?</b></p> <p>( ) Yes ( ) No ( ) I do not know</p> <p><i>If the answer is “Yes”, ask for a copy of the document. Otherwise, inquire whether the document could be shared under a confidentiality agreement.</i></p>
<p><b>10. Does your institution possess a formal norm establishing a systematic process for the execution of evaluation activities?</b></p> <p>( ) Yes ( ) No ( ) I do not know</p> <p><i>If the answer is “No” or “I do not know”, jump to question 12</i></p>
<p><b>11. Is the norm mentioned on question 10 a public document?</b></p> <p>( ) Yes ( ) No ( ) I do not know</p> <p><i>If the answer is “Yes”, ask for a copy of the document. Otherwise, inquire whether the document could be shared under a confidentiality agreement.</i></p>
<p><b>12. Does your institution possess formal documents regularly published with the results of the evaluation activities?</b></p> <p>( ) Yes ( ) No ( ) I do not know</p> <p><i>If the answer is “No” or “I do not know”, jump to question 14</i></p>
<p><b>13. Are the documents mentioned in question 12 public?</b></p> <p>( ) Yes ( ) No ( ) I do not know</p> <p><i>If the answer is “Yes”, ask for a copy of the documents. Otherwise, inquire whether the document could be shared under a confidentiality agreement.</i></p>
<p><b>14. Does your institution possess a formal norm establishing a systematic process for the review of evaluation results?</b></p> <p>( ) Yes ( ) No ( ) I do not know</p>

*If the answer is “No” or “I do not know”, jump to question 16*

**Section Two – Overview of norms related to evaluation activities (cont.)**

**15. Is the norm mentioned on question 14 a public document?**

- ☐ Yes
- ☐ No
- ☐ I do not know

*If the answer is “Yes”, ask for a copy of the documents. Otherwise, inquire whether the document could be shared under a confidentiality agreement.*

**16. Does your institution possess a formal norm establishing a systematic process for communicating evaluation findings to stakeholders?**

- ☐ Yes
- ☐ No
- ☐ I do not know

*If the answer is “No” or “I do not know”, jump to question 18*

**17. Is the norm mentioned on question 16 a public document?**

- ☐ Yes
- ☐ No
- ☐ I do not know

*If the answer is “Yes”, ask for a copy of the document. Otherwise, inquire whether the document could be shared under a confidentiality agreement.*

**Section Three – Overview organizational structure related to evaluation activities**

**18. Does your institution possess a specialized unit dedicated solely to evaluation activities?**

- ☐ Yes
- ☐ No, evaluation activities are performed for several different units
- ☐ No, evaluation activities are not performed
- ☐ I do not know

*If the answer is “No, evaluation activities are performed for several different units”, move to question 20. If the answer is “No, evaluation activities are not performed” or “I do not know”, move to question 22.*

**19. What kind of activities are under direct responsibility of this specialized unit (select all that apply)?**

- ☐ Registry of performance indicators
- ☐ Analysis of performance indicators
- ☐ Consolidation of internal datasets
- ☐ Execution of evaluation activities
- ☐ Hiring of external evaluations
- ☐ Execution of Audit and compliance activities
- ☐ Others – specify: \_\_\_\_\_

**21. Approximately how many employees are there in your organization?**

**Section Four – Overview of evaluation practices**

*This section has only open-ended questions, and, ideally, the answers should be recorded after a formal authorization of the employee.*

**22. What are the main objectives of the evaluation activities performed in your organization?**

*Some typical goals are organizational learning, performance improvement, accountability, and marketing. Try to provoke the employee to articulate the goals with organizational structure and norms.*

**23. What are the main methods applied by your institutions on evaluation activities?**

*The employee must be provoked to talk about the main techniques used to perform evaluation activities. The researcher must be able to identify whether the main causal assessment techniques are regularly used and how data-intensive is the work.*

**24. Do operational teams (those working directly with core-business activities of the organization) engage with evaluation activities? If yes, describe their role.**

*It is important to understand how much the operational teams have real responsibilities under the evaluation system.*

**25. What are the main consumers of the evaluation results?**

*Some typical answers would involve operational teams, marketing teams, high administration, control bodies, or other external stakeholders. Try to provoke the employee to articulate the consumers with the main objectives previously declared.*

## Appendix B: Description of the simulation exercises

These simulation exercises should be applied to a representative sample of medium and low-rank employees of the organization, encompassing workers from operational and support activities.

### **Simulation 1<sup>11</sup>:**

*This exercise starts with a general description of a theoretical policy/program that is closely related to the core business of the organization under evaluation. This description would include:*

- i) an explicit theory of change, with declared development outcomes;*
- ii) a statement that the policy/program still does not have any kind evidence-based analysis about its results;*
- iii) a statement that the dominant perception inside the institution is that the policy/program has been a success, with social benefits largely exceeding implementation costs; and*
- iv) a statement about the dominant perception about the program/policy outside the institution – this would be randomized, with 50% of the respondents dealing with a positive external perception and the others with a negative external perception.*

*After this general description, the employee would be required to picture himself as the manager in charge of the policy/program and to answer the following question.*

### **Question 1:**

*A team of evaluation experts has come to you with three evaluation proposals for the program/policy. Each one of these proposals applies different methods and datasets and, therefore, can be used for distinct purposes:*

- i) The first proposal can generate some information for positive advertising, helping to keep a favorable image of your institution;*
- ii) The second one is adequate for complying with external evaluation requirements, imposed by a national regulatory institution, but it won't serve to market the results;*
- iii) The third one will be able to implement a cost-effective analysis of the policy/program, but due to its high technicality, it will not be useful for advertising or for external evaluation requirements.*

**Considering this, which share of the total policy/program budget are you willing to allocate to evaluation activities?**

For the <b>first evaluation proposal</b> (select a number between 0% and 100%):	
For the <b>second evaluation proposal</b> (select a number between 0% and 100%):	
For the <b>third evaluation proposal</b> (select a number between 0% and 100%):	
Total directed for evaluation activities (the sum of the other fields):	

<sup>11</sup> The exercise is written for an application with public organizations. The adaption for an application within a commercial bank would include: i) changing references of policy/program to typical financial products; ii) changing mentions to development outcomes to profits.

**Simulation 2<sup>12</sup>:**

*This exercise starts with a general description of a theoretical policy/program. This description would be randomized, with 50% of the respondents facing a case that is closely related to the core business of the organization under evaluation, and the others dealing with a case that shown weak or no relation with it. This description would also include:*

- i) an explicit theory of change, with declared development outcomes;*
- ii) a statement that the dominant perception inside and outside the institution is that the policy/program has been a success, with social benefits largely exceeding implementation costs; and*
- iii) a statement that this dominant perception is not supported by any causal evaluation or another form of evidence-based analysis.*

*After this general description, the employee would be informed that the internal evaluation team has come with new evidence about the effectiveness of the policy/program. After a short explanation about the evaluation strategy, the employee would see a data presentation template inspired by Banuri et al. (2019). It would show observed values for a declared outcome variable on treatment and control groups, and the values would be calibrated to indicate a slightly negative impact of the policy/program.*

*After this problem presentation, 50% of employees facing a closely related case would be required to picture themselves as managers in charge of the policy/program. The other half would be required to put themselves in the position of a regular citizen. Under these different contexts, they would answer the following equivalent answers:*

**Question 1:**

**Considering the data presented by [your / the] evaluation unit, how [would you / should authorities] interpret the effectiveness of the policy/program [under your management / under analysis]:**

- ☐ Very disappointing
- ☐ Disappointing
- ☐ The evidence is inconclusive
- ☐ Positive
- ☐ Very positive

**Question 2: The understanding of the experts in the evaluation unit is that the policy/program is responsible for a relevant social welfare loss. For them, without a design change in the policy/program, the resource should be directed to other ends. [Your organization is / Public authorities are] now trying to find the best**

<sup>12</sup> The exercise would be slightly different for the 50% of commercial banks that will face a case close related with the core-business of the organization. For them, the dominant perception of “positive social impact” would be attached to a marketing strategy that promotes the financial product. Without this overall perception, the bank would face a profit reduction. The possible strategies would also be adapted. The decision to redesign the policy/program would be replaced by an option of redesign the marketing strategy, abandoning the narrative of positive social impact.

strategy to deal with these evaluation findings. How likely are you to support the following alternatives?

**2a. This is just one study, and as long as it is not published, the dominant perception about the policy/program will not change. There is a reason why so many people think the policy/program is a success, and [we / public authorities] should keep executing it.**

- ☐ Very likely
- ☐ Likely
- ☐ Indifferent
- ☐ Unlikely
- ☐ Very unlikely

**2b. This is just one study, and as long as is it not published, the dominant perception about the policy/program will not change. There is a reason why so many people think the policy/program is a success, and [we / public authorities] should keep executing it. Still, [we / public authorities] should work on a new evaluation to investigate further the actual impacts of the policy/program.**

- ☐ Very likely
- ☐ Likely
- ☐ Indifferent
- ☐ Unlikely
- ☐ Very unlikely

**2c. This is just one study, and as long as is it not published, the dominant perception about the policy/program will not change. Since so many people think the policy/program is a success, [we / public authorities] should keep executing it. Still, [we / public authorities] should start to redesign the program, trying to improve its effectiveness.**

- ☐ Very likely
- ☐ Likely
- ☐ Indifferent
- ☐ Unlikely
- ☐ Very unlikely

**2d. This is the first study about this policy/program, and as such, it must be publicly available. Still, since so many people think the policy/program is a success, [we / public authorities] should only publish it after [we / they] come up with a plan to redesign the program, trying to improve its effectiveness.**

- ☐ Very likely
- ☐ Likely
- ☐ Indifferent
- ☐ Unlikely
- ☐ Very unlikely



**2e. This is the first study about this policy/program, and as such, it must be publicly available. [We / Public authorities] should see how these findings will affect the internal and external perceptions about the policy/program before making any decisions related to the continuity of the policy/program.**

- ☐ Very likely
- ☐ Likely
- ☐ Indifferent
- ☐ Unlikely
- ☐ Very unlikely