EVALUATION BENCHMARK

Prepared for F3E by:

Josette Bockelie, Emma Ghariani, Kenley Jones, Alma Osorio, & Ieva Raudonyte

CONTENTS

I. Introduction

Why this report?

This report is the product of collaboration between F3E and student group project for the Master's course Scaling Innovation at Sciences Po Paris. F3E's request originated from the observation that while evaluation has very precisely defined terms when framed in the evaluation process, the overall objectives and understanding of evaluation by its stakeholders remain widely heterogeneous. This is partially due to the fact that there are a wide variety of stakeholders involved in the evaluation process, whose objectives vary and at times contradict one another. A lack of coordination among stakeholders causes the use and usefulness of evaluations to be constantly questioned. This report aims to create "a benchmark of the perception of evaluation by its stakeholders."

F3E partnered with Sciences Po to address a number of questions in order to provide a transversal analysis on the current issues of evaluation such as:

- Strategic issues:
    - What are the current objectives that evaluation addresses today?
    - What are the major positive changes brought by evaluations (at individual, organizational and societal level)?
    - What are the objectives that are not or not well addressed?
- Challenges:
    - What are the main challenges that the different stakeholders face?
    - What are the limits of current evaluation approaches?
- Recommendations:
    - What are the new evaluation approaches that are put in place or that could be explored?

The report initially sought to understand and create a worldwide benchmark of regional differences (cultural and linguistic ones) of current evaluation practices. We set out to better understand how various stakeholders view evaluation and how these views may vary across regions.

Furthermore, this report was commissioned by Charlotte Boisteau from the French organization F3E in the framework of the elaboration of the Barefoot Guide n°5 on Evaluation and Social Change. Thus, we approached our analysis with a thematic focus on this issue. Our analysis englobes previously raised questions often linking them to the idea of evaluation as a tool for social change. To combine the thematic and geographical approach commanded by F3E, we coordinated our work around a matrix of thematic questions, that respond to the initial issues that aim to respond to the original inquires presented, crossed with geographic features, so as to be reusable and expandable in future research for the Barefoot Guide.

An evolving methodology

Scoping meeting

During the first scoping meeting, we decided to reduce the scope of the study. Given the scope and the demanding nature of the project, we have decided to focus on evaluators and how they interact with other stakeholders. Evaluators provide a useful entryway into the topic of evaluation, because of their experience working on a variety of projects and in different regional contexts. Also in an attempt to more narrowly frame our research, we agreed to focus exclusively on external, ex-post evaluation. It should be noted, however, that evaluators are merely one of the stakeholders involved in the evaluation process. Therefore, this research should be viewed as the preliminary step in a longer process of learning how donors, implementing organizations, and beneficiaries feel about the evaluation process.

Desk research & first outline

In the first step of our work, we focused on desk research, both on issues we were expected to deal with (learning, accountability, methodologies, social change…) and on how geographic differences could impact evaluation methods and results. Our report includes a brief overview of the existing evaluation literature. Nonetheless, we found some peer reviewed academic articles dealing with their influence of geographical and linguistic characteristics on the evaluation process.

Based off our literature review, we elaborated broad interview matrix. This broad interview matrix (see annex) includes different questions that should be investigated: they are grouped according to the three issues (discussed in more detail in section II) and three main actors that we have identified*:* donors, evaluators and implementing organizations. We created a diagram outlining the different interactions between stakeholders (see annex for more details).

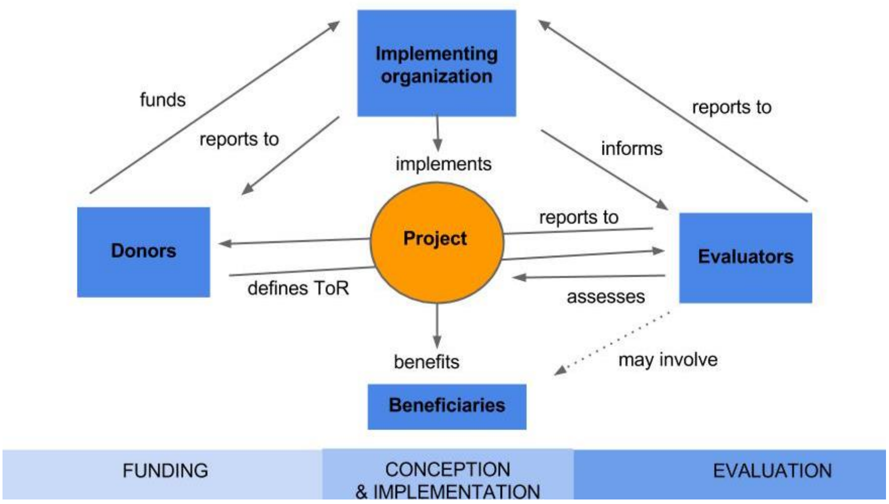


Figure 1: Stakeholder interactions

We considered these to be key stakeholders; however, we acknowledge that some others could also be included in the future analysis. Furthermore, it is necessary to clearly underline that we acknowledge that these groups inside themselves are very heterogeneous and this should be taken into account throughout the analysis. For example, national agencies might be at the same time donors, operators of the project and/or evaluation. Big donors might participate in the evaluation and not solely ask for it. However, in this work, donors will mostly be considered as agents asking for evaluation for the projects that they funded; evaluators as agents conducting evaluations, and implementing organizations (often referred to only as *organizations*) as the ones that are actually working in the field and whose work is often evaluated. Nevertheless, this distinction rarely manifests in clearly definitive categories. Furthermore, it is necessary to define what the term "beneficiary" denotes. Florent Bedecarrats from the AFD, for example, mentions that the beneficiaries play a central role in the evaluation process of development projects, followed by a specification that for the AFD the term "beneficiary" refers to the actor with whom they are working in the project country – thus the government or a minster. In terms of the local population or community that actually *benefits* from a project, the AFD uses the term "final beneficiaries." While the government or related ministry certainly benefits from development projects, we consider the true "beneficiaries" the community and population that would directly benefit from a project or program. Thus, the term is employed in this sense throughout the report.

As previously mentioned, we focus our work on evaluators, as they are the logical entry point. From our broad interview matrix, we used the questions pertinent to evaluators to create an interview guide common amongst all our interviewees (with the exception of the question related to regions, which was included halfway through the process) (see annex). We envisage the remaining two sections of the broad interview matrix, regarding donors and implementing organizations, to serve as a blueprint to create interview guides for future use to gain a more holistic understanding of the interaction of stakeholders involved in the evaluation process.

Interviews

We conducted qualitative assessment of evaluation based on fifteen interviews. A majority of them were conducted with the evaluators themselves or people closely related to implementation of evaluation and with an advisory platform working on evaluation. The contacts were obtained in part thanks to Charlotte Boisteau and in other part because of personal contacts or through the evaluators themselves. The interviews aimed at gaining a better understanding of how evaluators perceive and feel about the evaluation on the ground. As expected, the results changed our scope again.

Due to the background of our interviewees, our assessment will focus mostly on evaluation of development projects. Although our literature review focused on ex-post external evaluation practices, many of our interviewees, argued that there is a need for more reflective, inward looking evaluations. Interviewees suggested that, in the future, the development community would become increasingly reliant on internal evaluations. Moreover, some of the interviewees underlined the importance of monitoring and midterm evaluations. Ignoring these comments would have weakened the substance of our work; therefore, we expanded our original paradigm to include some perspectives on internal evaluations, monitoring and mid-term evaluations in our assessment portion of the report.

Interview analysis

The interview analysis enabled us to draw connections between the interviews so as to see transversal issues and trends in current evaluation practices. As we seek to provide an unbiased perspective on evaluation as defined during our scoping meeting, we attempt to keep a neutral stance on the opinions emitted by our interviewees; we deem all opinions equally valuable and attempt to present conflicting perspectives when they occurred so as to inspire debate not only on the current purpose of evaluation but also its future.

We compiled the interviews into three broad issues (defining evaluation, learning, and innovation). We then subdivided these themes into more specific issues. Part way through the interview process, we added an additional question targeted at understanding geographical and linguistic differences in evaluation practices. Ideally, we sought to deduce these aspects from the interviewees' answers without asking directly, which proved extremely difficult. Since this was a perspective that we had to introduce rather than our interviewees noting it as a major theme in evaluation, we consequently decided to reduce the part for geographical analysis. Even with interviews with individuals working in evaluation across the globe, we lacked sufficient data to attain a meaningful conclusion on regional differences. We chose to replace a detailed discussion of regional differences with a critical analysis of why this did not appear in the interviews and why other criteria plays a more influential role in determining evaluation culture and norms.

A first draft was submitted to F3E for remarks that enabled us to rework the report to provide a final version that provides as holistic perspective as possible on the current trends in evaluation and the potential to use evaluation as a tool for social change. We aimed to produce a report that would be as operational as possible for both the elaboration of the next Barefoot Guide and F3E. We hope that this report will act as a stepping-stone to pursue further research on the various stakeholders involved in evaluation and their interactions.

II. Rationale for Selecting Issues

Summary of research

While reviewing the existing literature on evaluation, we selected three key issues on which to focus. This helped us narrow our perspective in a field that is widely researched and allowed us to study three key aspects of evaluation in greater detail to give a more profound understanding to the topic. First, we emphasize the need to look at how different stakeholders define evaluation. Next, we explore the relationship between accountability and learning in the evaluation process. Finally, we address innovative practices that could be observed in evaluation.

Issue 1: Defining Evaluation

Defining evaluation in this benchmark serves two purposes. First, it lends clarity to the project and provides for a mutual understanding of terminology. Secondly, it allows us to identify any possible misconceptions regarding the purpose of evaluation. According to the World Bank, evaluation "is the systematic and objective assessment of an on-going or completed project, program, or policy, and its design, implementation and results. The aim is to determine the relevance and fulfillment of objectives, development efficiency, effectiveness, impact, and sustainability." While this definition succinctly captures what evaluation *is*, there remains a need to identify how various stakeholders *perceive* evaluation. Previous studies suggest that individuals negatively associate evaluation with auditing or rules imposed from the top-down (Conlin & Stirrat, 2008; Estrella & Gaventa, 1998; Picciotto, 2007). Professor Des Gasper of the International Institute of Social Sciences (ISS) at the Erasmus University Rotterdam argues that as evaluation has become more prevalent in the development community, however, it has also become more holistic; it aims not only to assess how money was spent, but also to contribute to accountability and learning (Gasper, 2000).

Although the aims of evaluation have been laid out in various studies, it is important to ask questions to see how evaluation actors perceive those aims. Is there general consensus amongst actors and stakeholders about what evaluation is meant to achieve? Do different stakeholders have different goals when it comes to evaluation? Clarifying how various stakeholders define evaluation is the first step in understanding the implications that evaluation can have for learning and social change.

Issue 2: Evaluation and Learning

With the adoption of the Millennium Development Goals in 2000, the international community has increased its focus on long-term development, rather than short-term aid. This shift in focus also alters the role that evaluation plays. It is no longer sufficient to track how projects and programs spend aid money; instead, organizations need to know how effective their programs have been in creating behavioral change (Earl, Smutylo, and Carden, 2001). The fact that many countries did not meet their MDG goals has only heightened the awareness that assessing the extent to which programs achieve progress in reaching development goals is paramount. This increase in the desire to measure results is clear in the new Sustainable Development Goals (SDGs). Goal 17.18 and 17.19, explicitly address the need to gather data for results-based management. The SDGs focus on the accountability aspect of evaluation rather than the learning aspect. Gathering statistics, however, will not be enough to promote development; the development community will also need to assess their shortcomings in the MDGs. Learning from past mistakes represents a crucial aspect of progress. Therefore, we are interested in assessing the degree to which evaluation plays a role in the learning process.

The desire to incorporate learning into the evaluation process is also evident on international development job boards such as Devex. An increasing number of organizations are looking to employ monitoring and evaluation practitioners who will focus on building knowledge as well as measuring impact (Rogers, 2015). Knowledge and impact, however, are two very different things. Measuring impact falls in line with the traditional role that evaluation has in addressing accountability. Building knowledge, on the other hand, focuses on evaluation as a tool for learning. This also begs the question as to which type of evaluation organizations prefer. While some organizations hire external contractors, others are building internal evaluation departments. To further assess evaluation as a medium for exchanging knowledge, it will be necessary to consider how various stakeholders interpret the relationship between evaluation and learning.

Issue 3: Innovation

At the 2013 First International Conference on National Evaluation Capacities, EvalPartners, a global network of evaluation actors, declared that 2015 was the International Year of Evaluation. Designating a year devoted to evaluation aims to promote innovation in capacity building, communication, and strategic partnerships. Therefore, it is important to identify current trends in evaluation, areas for improvement, and innovative practices that will help evaluation stakeholders achieve their goals.

As evaluation is currently seen as means to measure outputs, outcomes, and impact, the methods used fall in line with these goals. On the other hand, evaluation as a tool for social change implies a change in methodology and practices. As evaluation becomes more holistic, it will seek not only to measure changing behaviors within a society, but also to produce positive social change; in doing so, evaluation will need to use innovative methods and tools. Social change implies a set of evaluation questions, judgment criteria and indicators that are radically different from existing ones. In order to collect information, evaluators will need to rely more heavily on innovation. Innovation in our work will be broadly defined as an adoption of relatively new techniques or tools in order to address new challenges, which ask evaluation to become more holistic and encompass aspects such as learning or social change.

Regional Overview: Literature

According to the American Evaluation Association (AEA), "Evaluations cannot be culture free. Those who engage in evaluation do so from perspectives that reflect their values, their ways of viewing the world, and their culture. Culture shapes the ways in which evaluation questions are conceptualized, which in turn influence what data are collected, how the data will be collected and analyzed, and how data are interpreted" (*Public Statement on Cultural Competence in Evaluation*, 2011). Since culture is vital in shaping the way individuals view the world, we sought to identify regional differences in how evaluations are perceived and realized.

We will also consider that some regions have a greater influence than others. In formulating this project, we spoke with James Taylor of the Community Development Resource Association (CDRA) in South Africa. He highlighted the role that power dynamics play in defining development agendas and evaluation practices. According to Taylor, all too often, evaluations become a tool that wealthy nations use to assess the work of poorer nations. Thus, instead of measuring the true impact of a development project, evaluations leave individuals feeling judged and disempowered. Keeping in mind this shortcoming, we will consider not only the culture in which the evaluation takes place, but also who defines the culture of the evaluation itself. In doing so, we will incorporate a broad interpretation of culture that reflects both geographical traditions, as well as methodological and thematic norms related to culture.

1. Anglophone Europe – USA – Pacific

First of all, it is essential to underline that it was in North America where evaluation emerged as a distinct area of professional practice in the 1950s and 1960s. This explains why the evaluation tradition is so strong in this region. This also characterizes other Anglo-Saxon and Nordic countries that might be considered as more "evaluation oriented". Even though evaluation practices among these countries might vary and deserve a more profound and detailed analysis, some general trends and characteristics exist. The literature underlines many similarities between the evaluation culture in the United Kingdom and the United States. These regions tend to underline the accountability dimension. At its conception, the main justification for evaluation was a need for accountability.

A second important characteristic is the positivist nature of evaluation in this region. The US social sciences, in general, can be considered more embedded in the positivist school than it is the case in some of the European countries. Positivism "has at its heart the belief that it is possible to obtain objective knowledge through observation. Different people applying the same observation instruments should obtain the same findings. Positivist traditions aim to discover regularities and laws (as in the natural sciences). Explanations rest on the aggregation of individual elements and their behaviors and interactions." (*EVALSED: The resource for the evaluation of Socio-Economic Development, September 2013, p. 18).* These elements are highly present in evaluation practices of the region.

Moreover, as previously mentioned, Scandinavian countries are more evaluation-oriented than some other European countries (especially the ones in the South). Nordic countries were among the first ones to introduce evaluation practices, which are currently very strongly rooted in their public policy making. They are keen to discuss evaluation results and were active in promoting an evaluation culture in the European Union context.

2. Francophone Europe – Quebec

A. France

Public Policy evaluation in Europe began with the UK and the Scandinavian countries in the 1960's; it has since spread to become a norm all throughout Europe in large part due to the European Union in the 1990's. The EU required systematic evaluation of projects and programs in order to receive financing for national development projects.

France first introduced evaluation as an aspect of their public decision making process with *rationalization des choix budgétaires* (RCB-Rationalizing budgetary choices) in 1968, which encouraged ministers to dedicate a portion of their budget to *ex-ante* evaluations (la documentation française, 2004). This policy, however, was abandoned in 1984 due to increased budgetary restraints. In 1990, a decree was passed relative to public policies, which created a means to complete interministry evaluation with the *Conseil Scientifique de l'Evaluation* (CSE-Scientific Evaluation Counsel). In 1998 the CSE became the *Conseil national de l'évaluation* (National Counsel of Evaluation). While the formal application and laws englobing evaluation in France are relatively new, the notion of evaluation goes back to the foundation of the first Republic with the *Déclaration des droits de l'homme et du citoyen* (1789) (Declaration of the Rights of Man and the Citizen), in which the preamble states "*la société a le droit de demander compte à tout agent public de son administration.*" (Society has the right to request a report from all public agencies and the administration).

France's approach to evaluation of public policies can traditionally be divided into four categories, each with a distinct usage (CSE, 1996). (1) Evaluation of impact to see the effects of a project or program on society. (2) Evaluation of the process to analyze the way in which a project or program was launched and whether that contributed to the projects' success or failure. (3) The evaluation may be directed to decision makers to aid in them in their process; or, (4) as a tool of communal learning so that agents better understand their role and objectives, which can clarify points that are not useful. The French perspective has traditionally focused on evaluation as a means to ensure that public policies, projects and programs are the most efficient (uses resources wisely) and effective (responds to the initial need) as possible. Evaluation of development projects and programs has increasingly focused on learning in addition to accountability.

France has been mentioned as more open to innovation in public policy evaluation because French evaluations already implicate everyone in the steering committee. Belgium, on the other hand, has been criticized as having a weaker evaluation culture, where government takes a defensive stance to evaluation (OECD, 2010). This translates into a steering committee that seeks systematically avoid conflict. One recommendation received by the OECD Peer Review in 2010 was to include more stakeholders; by creating a dialogue between evaluation actors, Belgium could create a more comprehensive and inclusive evaluation culture and enhance synergies. An additional critique relates to learning from evaluations: the evaluation process ought to better incorporate lessons from evaluations as a means to strengthen evaluation culture build capacities (OECD, 2010). These recommendations follow the French association F3E's methodology of evaluation in international development projects. F3E promotes and offers tools for the capitalization of experiences in a collective manner that allows for both individual and organizational learning (F3E, 2014). F3E's tools for evaluation also include guidelines for a "change oriented approach" to evaluation. The goal of evaluation is to accompany a change in society that includes capacity building of the involved actors, increase awareness of citizens through education, and lobbying (F3E, 2014b). It requires a reconsideration of the final goals of a project away from technical aspects of an intervention and toward long-term goals. This changes the linear perspective of evaluation; evaluation can no longer be seen in simple cause and effect terms (F3E, 2014b)

B. Quebec

In opposition to the French model, which respects the traditional methods of evaluation defined by the DAC, the Canadian model includes innovation as an integral aspect of evaluation and the future of evaluation as a tool for social change, according to Shore (2008) in a report for the CRDI. For instance, the CRDI held a series of conferences in 2008 relative to Innovation in Evaluation. The CRDI also utilizes *cartographie des incidences* (outcome mapping), an evaluation method based on changes in comportment, the relationships and activities of individuals in the organizations in which they work as opposed to merely the external products of a project or program (Shore, 2008). With social innovation making up an important part of the economy and, organizations and evaluators recognize the increasing value that innovation can play in evaluation (Shore, 2008). For example, a central aspect of the evaluative questions is innovative responses and solutions to present problems or difficulties (Duran & Monnier, 1992).

3. OECD perspectives

While the OECD regroups a variety of countries, with member countries coming from nearly every region, a discussion of the OECD's key additions to evaluation culture seems necessary. In 1991, the OECD Development Assistance Committee (DAC) defined Evaluation Principles for Development. They defined five criteria that each evaluation ought to respond to: relevance, efficiency, effectiveness, impact and sustainability. With the Paris Declaration in 2005, evaluation standards and principles were established for development agencies and partner countries with a goal of enhancing aid delivery and management. The recommendations of the Paris Declaration include "local ownership of development strategies; alignment with national development strategies, the harmonization of development actions; the evaluation of results and mutual accountability and transparency" (AFD, 2014). Nonetheless, the Paris Declaration represented a heavily institution-oriented conception of evaluation. It was not until 2008 with the Accra Agenda for Action that civil society organizations' complementary role in evaluation was recognized and valorized. Around this time, a paradigm shift occurred; evaluation became increasingly focused on development effectiveness and less focused on its original mandate as a tool to measure aid effectiveness. Charlotte Boisteau notes that "today, methodologies are evolving towards a more holistic and qualitative assessment and an inclusive pedagogical learning-oriented evaluative process, focused on the points of view and participation of actors" (AFD, 2014, 16). As a development tool, evaluation plays a different role and takes a different approach than it did as a measure of accountability of public aid policies. This shift in focus is what this report attempts to understand and analyze from the perspective of how evaluation can be used as a *tool* in development and social change that accompanies it.

4. Asia and Africa

The majority of development aid is sent by western countries and received by the global South; this creates a context in which Western European and North American countries largely determine the development agenda. Scholars have long noted the problems arising from the fact that partner countries have little say in determining their own development agenda (Brohman, 1995). In the past decade, following the 2005 Paris Declaration on Aid Effectiveness, there has been a heightened awareness of the need for local ownership of the development projects. Despite the acknowledgement that the development process should be owned by developing countries, however it continues to be largely influenced by outsiders. Actors in many Western nations are aware of the problems arising from the lack of grassroots involvement in development projects. For example, The Modernizing Foreign Assistance Network (MFNA), a coalition of citizens and international development practitioners that advocates for more effective U.S. foreign assistance, highlighted local ownership as one of the key targets for improving aid; yet, it remains unclear how effective the international development community has been in realizing this goal (The Way Forward: A Reform Agenda for 2014 and Beyond). According to James Taylor, evaluation in particular remains heavily influenced by wealthy countries. In other words, there is a rift between the stated aims of the international community to allow local ownership, and the reality on the ground. The following sections focus on evaluation trends in Africa and Asia.

A. Africa

In 1999, the African Evaluation Association (AfrEA) was founded in response to the need to promote an evaluation approach specific to the needs of African countries. This and other associations of African evaluators (such as Senegalese Evaluation Association) have tried to build up on evaluation guidelines that are "sensitive to the African context and (…) in line with international practice, standards and norms for evaluations," (African Evaluation

Association, 2007, p. 4) Interestingly, the AfrEA's *Guidelines - Standards and Norms* sheds light on different opinions that African evaluation consultants have on the adaptation of international evaluation norms to African values. Hence the reports accounts that there was a "first perspective was that it is acceptable to adopt an international model that had sufficient sensitivity to the African context. (...) The second perspective was that it was unacceptable to impose an externally developed set of standards on Africa. (...) The third perspective was that the appropriate procedure would be to test the US PES in field conditions in Africa in order to determine their suitability and to identify any modifications that might be required on a pan-African basis" (African Evaluation Association, 2007). In the end for Oumoul Kharyi Ba Tall, an evaluator with OKT-Consult in Mauritania, these different opinions reflect debates that more generally occur within the arena of international evaluation networks. This attempt to build African evaluation guidelines has produced a methodology including 35 standards divided into 4 major principles:

1. Utility principle: for produced information and expected and provided results.
2. Feasibility for realism: cautiousness and efficiency.
3. Respect of ethics: respect of legal and ethical rules.
4. Precision and quality: for a relevant methodology related to the goal and the subject matter of the evaluation

An overarching theme in the AfrEA's approach to evaluation is the need to incorporate stakeholders in the process. In order to make evaluations more localized and context specific, it is important to engage with project beneficiaries throughout the evaluation process (African Evaluation Guidelines, 2007).

B. Asia

The Asian Development Bank (ADB) is one of the key actors in defining evaluation practices in Asia. The central goal of evaluation at the ADB is to move past measuring project outputs and instead focus on project outcomes (Thomas and Luo, 2012). It advocates an expansive results-based approach to evaluation. As opposed to monitoring expenditures, evaluating for results "increases the importance of looking beyond the stated objectives to find synergies and cross-cutting solutions to complex development issues" (Wang, 2014). The ADB clearly defines learning a key evaluation objective.

The Participatory Research in Asia (PRIA) center argues that international organizations tend to overlook the importance of learning in the evaluation process. PRIA has been active in promoting participatory monitoring and evaluation (PM&E) as a means to increase learning. The aim of PM&E is to shift the focus from results to learning by involving program beneficiaries in the evaluation process (PRIA, 2015).

5. Hispanophone Latin America

Over the last decades, Latin America has been working on strengthening the backbone of their evaluation capacity. According to a report released by the Programme for Strengthening of the Regional Capacity for Monitoring and Evaluation of Rural Poverty-Alleviation Projects in Latin America and the Caribbean, the main focus of the region has been the institutionalization of project monitoring and evaluation. They have conducted surveys from 26 different projects in the region, and obtained responses from 24 different directors. Key aspects of their conclusions are that monitoring and evaluation systems were put in place in all projects surveyed. Additionally, they show a steady increase in the implementation of their evaluation system, which they attribute to higher quality interaction and participation between stakeholders. Furthermore, they have also noted more presence of technical personnel in the projects in question, which they interpret as a factor that can exemplify the institutionalization of monitoring and evaluation in the region. On the other hand, they do come across information that confirms the relatively low budgets for monitoring and evaluation systems. In fact, some projects do not have funds allocated at all for such activities, and coordinate rather within their organization to find the means to carry on monitoring and evaluation systems.

The Office of Evaluation and Oversight falls within the charter of the Inter-American Development Bank (IADB). They are quite active as they evaluate all the IADB projects in the region. They focus heavily on project evaluation comparisons. Their themes vary, but they focus mainly on environment, citizenship, migration, and natural resources projects. One of their main limitations is that their impact evaluations are carried out exclusively on projects undertaken by their investment partners. Given their presence in virtually the entire region. Their methodology and scope could give monitoring and evaluation in the region a more homogeneous language.

A. Central America

The German Foreign Ministry for Economic Cooperation and Development commissioned a study on the Evaluation Capacity Development in Latin America. Their findings on the Central region shed a different light in the prevalence of evaluation system in the region. They acknowledge that the region still comes at a deficit in terms of refining funding for monitoring and evaluation systems put in place. They blame, for the most part, a lack of political will to set aside public funds to finance these systems. They also point out lack of technical knowledge to carry on with proper evaluations that meet international standards. Finally, they insist on the need to create a common language for the concept of evaluation in the region. These findings are surely very specific to the criteria the Ministry is utilizing to frame evaluation in the region; and can vary depending on the approach, i.e., stakeholders' involved.

III. Interview Assessment: Evaluation, Learning, & Innovation

This section constitutes the heart of our research work. It provides a transversal analysis of all of our interviews regarding the three major issues identified at the beginning of our work. Our assessment seeks to give a synthesis of the ideas expressed by the interviewees: analyzing, contrasting and complementing aspects. This part seeks to provide a comprehensive analysis of what interviewees have expressed during our interviews.

Issue one: Defining Evaluation

1. Donor defined process

According to the majority of our interviewees, donors play the primary role in defining the evaluation process. While many interviewees stated that organizations could learn a great deal by informally assessing their projects, donors drive the formal evaluation process. James Taylor of the CDRA pointed out that a donor-driven evaluation process could lead organizations to feel disempowered; in this case, donor-driven evaluations cause negative relationships between organizations and evaluators.

Patrice Dufour explains that while the curiosity to better understand why development projects succeed or fail motivates him personally as an evaluator, the process may not reflect this curiosity sufficiently due to donor's strong influence in shaping the process. Regardless of the personal motivations of evaluators or organizations, evaluations typically take place because donors are asking for them. Furthermore, Dufour is skeptical about what donors intend to learn from evaluations. The goal of a donor, such as the World Bank where Dufour previously worked as part of the Independent Evaluation Group (IEG), is to approve loans or grants. Therefore, evaluations sometimes become just an afterthought. They are a task that donors need to complete but not something in which the donor is fully invested.

According to Emilie Aberlen of the Agence Française de Développement (AFD), evaluations are largely donor driven under the perspective of "New Public Management": the flow of aid needs to be accountable to the people that gave the money and the project needs to be accountable for its results. Donor countries spend taxpayer dollars to fund development projects; therefore, they need to hold themselves accountable to their citizens by evaluating the impact of the money they spent in a foreign country. Florent Bedecarrats of the AFD and Jan Van Ongevalle of Katholieke Universiteit Leuven (KU Leuven) reiterate this point, saying that donors ask for results-based management and the fulfillment of the Organization for Economic Cooperation and Development's Development Co-operation Directorate (OECD/DAC) criteria. According to DAC criteria, development projects should be assessed on their relevance, effectiveness, efficiency, impact, and sustainability. Some interviewees add, however, that donors are not always the sole determiners of the evaluation. Some donors leave space for the implementing organizations (the ones that prepare the ToRs of the project and are currently working on the project implementation) to add their questions and concerns to the evaluation. They may raise certain questions that could be particularly relevant in the framework of their present or future projects and in this way these organizations can also participate in defining evaluation. Moreover, the evaluator also plays an important role in determining the approach to the evaluation. One of our interviewees explains that even when donors are responsible for setting the terms of the evaluation, the quality of the evaluation largely still depends on the individual evaluator. The same interviewee believes that while the donor gives important guidelines and sets the framework, in the end the evaluator also contributes to determine the main goal.

In recognizing the shortcomings of donor-driven external evaluations, some interviewees suggested that organizations might move more towards internal evaluations. Sue Soal of the CDRA argued that evaluations could be especially useful when they allow organizations to learn; for learning to occur, however, the evaluation report should not belong to the evaluator or to the donor; it must belong to the evaluated organization. If the organization has ownership of the evaluation, they could adapt it in ways that allow them to improve their practices. Although Soal believes evaluation has the potential to be a powerful tool for social change, she states that conventional evaluations are often a wasted opportunity in that they measure results rather than empower organizations to make meaningful, reflective changes. When donors define the entire evaluation process, evaluations do not always lead to findings that can help organizations improve their practices. Due to this problem, some evaluators predict that a shift in the evaluation field away from external donor-driven evaluations may occur. In their place, one might find more organizations using internal evaluations to assess their impact without donor constraints.

2. Links between monitoring and evaluation

The central aim of our work sought to analyze evaluation. *A priori*, we excluded monitoring from our analysis because of a clear distinction between monitoring and evaluation made by a large body of literature. Nonetheless, interviewees mention monitoring and its link to evaluation several times, often implicating a number of interesting ideas. Thus, we decided to integrate the link between monitoring and evaluation in our analysis.

Even though interviewees acknowledge a difference between the two concepts, a few of them underlined just how closely monitoring and evaluation are related (Patrice Dufour, Jan Van Ongevalle, Mary Hall). This relation is especially strong during the implementation phase of the project when monitoring and mid-term evaluations take place. Jan Van Ongevalle underlines that evaluation is essential during the implementation of the project in order to make adjustments: in this case, evaluation and monitoring need to go hand in hand. According to Van Ongevalle, these two activities must take place during the implementation of the project because otherwise people will not take time to collect information and to analyze the direct impact their activities are having. Project teams will tend to wait for someone to come after the project and complete an ex-post evaluation. This type of methodology, however, leads to the loss of very important information. Patrice Dufour also agrees that more evaluation should occur throughout the process of the project. All these ideas are closely related to the perception of evaluation as a learning tool that helps to learn directly from the possible weaknesses of the project not only in order to improve the ongoing project but also to ameliorate the future ones. A learning aspect of the evaluation will be addressed separately further in the work.

While ex-post evaluation comes later in the project cycle and provides a deeper and more nuanced analysis, monitoring can be more pragmatic as it is more concise. Mary Hall underlines that monitoring in some cases provides the quickest and most accessible way for project managers to get a direct feedback. The system was designed to identify lessons learned – in particular the grading system was to identify where performance was particularly good more generic lessons/good practices might be forthcoming or where there were serious shortcoming to be avoided in the future.

Almost all interviewees agree that monitoring and evaluation systems need to be built in the project from the very beginning. Some interviewees view this as one of the current shortcomings of evaluation. While the inclusion of monitoring and evaluation from the beginning of a project has become more prevalent, it is not always the case.

The strong connection between monitoring and evaluation and its implementation from the beginning of the project can not only help in readjustments when necessary but could also help induce social change, as underlined by some interviewees. When evaluation and monitoring are taking place during the project they might approach different parties that would sit together to look at what activities are being implemented and what are the direct impacts. This might lead different stakeholders to rethink what is happening which is a powerful motor for a social change, as mentioned by Jan Van Ongevalle. Moreover, Mary Hall, a monitor for the EU, highlights that if evaluation were to be used as a tool for social change, about which she expressed skepticism[1], the external ex-post evaluation process is almost powerless to fulfill this mandate within an individual project: thus, midterm evaluation going in hand with monitoring could be more powerful.

However, some interviewees were more skeptical about monitoring. Sue Soal underlined that evaluation should be reflective learning process rather than something forcing guidelines on individuals – which she believes monitoring does. Moreover, Ian Hopwood, former Chief of Evaluation at UNICEF (1996-2000), mentioned that some of the monitoring practices that people place in the same category as evaluation do not have anything to do with evaluation, or with the aims of the organization. For example, donors may ask organizations to monitor certain results, which do not in fact reflect the effectiveness or impact of the evaluation. Unlike some of the other interviewees, Soal and Hopwood draw a line between monitoring and evaluating the long-term impact of development work. These differing opinions suggest that relationship between monitoring and evaluation needs to be studied in more detail. On one hand, monitoring

1 This perspective will be further nuanced in the section on evaluation as a tool for social change.

provides data that is fundamental to an evaluator being able to do his or job. On the other hand, excessive monitoring can also be a useless exercise that waste's an organization's time and resources. Therefore, improving evaluation may very well require improving monitoring practices as well.

However, it is crucial to underline that different organizations might define and use M&E differently so when one person is talking about one or other they may not be referring to the same activities as another person. Semantics cause great confusion in the M&E world which should be taken into consideration.

Issue two: Learning and Evaluation

3. The importance of learning

Donors often use external evaluation consultants to measure the social impact of a program as a guarantee of external validity of the results from the programs that they fund. Most of the interviewees in our study strongly emphasized learning as an important outcome of the evaluation process; however, they were skeptical about the extent to which this goal is realized. There is also a strong debate among international networks and even some donors to use evaluations as a way to create and sustain learning in organizations.

For all of our interviewees, learning is a desirable outcome of the evaluation process. They each underline the potential for learning. According to Ian Hopwood, this space for learning is due to the multidisciplinary and multi-stakeholder nature of the evaluation that aims at confronting visions. Mary Hall points out that evaluations can foster learning either for future projects in the case of ex-post evaluations, or also for adjusting the program while it is implemented. Proactive adjustments can improve an ongoing project, adds Jan Van Ongevalle.

Interviewees highlighted several reasons why evaluations rarely fulfill the goal of learning. First, evaluation is too often seen as a negative judgmental process similar to audit. Evaluation is neither an audit, nor an inspection and it is important not to present it as such. One cannot just arrive and begin criticizing people because doing so would activate individuals' defensive mechanisms and the process would be blocked. Furthermore, the elements of accountability included in the evaluation process make learning more difficult. James Taylor, former director of the South African organization CDRA sees mutual accountability as a way to strike this balance and foster learning. The question is "Who is accountable to whom?... There is no mutual accountability." While implementing organizations have to be accountable to donors, donors are not accountable to implementing organizations in terms of disseminating the information of the evaluation or other lessons learnt which impedes the learning process of the organization (even though often those who agreed to contribute to the project indicate how they have done so or if not why not). Half of our interviewees evoke this problem of reciprocity, which hinders the learning process. They agree that it would be necessary to promote a participatory approach to learning, focused on both individual and organizational learning. The other interviewees understand learning as something that concerns mostly the donors: what programs should they fund and how should those programs be designed. Although this may not always be the case in reality, they see evaluation as a way to improve programs, or public policies in a more holistic way.

Nonetheless, the effective learning of the donors, implementers and all other stakeholders including final beneficiaries does not always occur. Too often, the evaluators learn more than the implementing organizations or even the donors. Although evaluation defines learning as a desirable goal to achieve, effective learning appears quite difficult to realize. The hurdles to learning can be divided into three broad categories related to the focus of evaluation.

A. Donor defined ToR

While consulting evaluators may learn something, it is rarely useful information for the implementing organization; the terms of reference do not include what the organization needs to learn, but rather what the donor wants to know. James Taylor emphasized on this point, saying that NGOs "need to learn what is really valuable. NGOs say that donors are not interested in learning about their most important work. Instead they just want to learn about the things they can count and measure. It is important to learn how people feel about evaluation." This underlines the conflict between the quantitative and qualitative evaluation approaches. Quantitative approaches rely on hard data and measureable results, but sometimes the most important work of an organization cannot be quantified in numbers. This is why Sue Soal suggests that learning will be improved when evaluations are more reflective and introspective; qualitative evaluations take into consideration the full scope of the organization's actions, rather than just those that can be quantified.

B. Secrecy as an impediment to learning

As evaluations are political processes and can involve large financial grants[2], a certain amount of secrecy may surround evaluations. Independent evaluation consultant and former director of the Atlantic foundations John Healy underlined in his interview that the results of the evaluations could or could not support learning but it depends on the aim of the evaluation. He noted the example of the Atlantic foundation, which uses some evaluations to make difficult decisions on whether to keep or terminate program. In this case, the results of evaluation would not be made public until after the decision of the administration board; thus, the evaluation results could not be widely debated. More generally, Pablo Rodriguez-Bilella, a consulting sociologist with EvalPartners, points out: "there are many restrictions in various cases regarding findings. This can be problematic, because transparency, specially related to public programs, can lead to results being stored and never looked at again. Anonymity, can lead to a roadblock on learning from the evaluation." This suggests that there is a need for increased coordination and communication between various evaluation stakeholders if the goal of learning is to be realized.

2 According to Vinod Thomas of the Asian Development Bank (ADB), impact evaluations cost $500 000 on average (Piccio, 2014).

Organizations such as F3E could play a role in reducing the secrecy surrounding evaluations by working with all those engaged in the evaluation process to build international solidarity and decentralized cooperation.

C. A lack of systematic follow-up

Both donors and organizations ought to have a more systematic follow-up of evaluation results. Our interviewees emphasized the need for consultants to follow-up the evaluation results and recommendations; the integration of results into organizations' everyday practices is unequal, and in some cases nonexistent. One reason for not integrating the results of an evaluation could be that the evaluation itself is not viewed as legitimate, especially if it has been mandated by donors, rather than generated by a desire for self-learning within the organization. Most noted a rupture of communication between the evaluator and the other stakeholders after the report is delivered. As follow-up is costly, it is usually only considered by international organizations or organizations with bigger budgets. Also, most interviewees note that the reports are too frequently stored away and never looked at again due to time constraints. Ian Hopwood explains that in the development field in general, there can be a "silo problem", meaning that certain departments or sectors are not willing to share information with others. Ideally, because evaluation is so interdisciplinary it can play a role of "breaking down silos" and ensuring information is shared.

There is a bias in how we understand effective learning because the learning process is not documented itself. Interestingly, Ian Hopwood highlights a problem of "organizational amnesia" – organizations do not learn from past projects or mistakes. There can appear to be a lack of learning because what the individuals actually learned throughout the project and evaluation is never documented. For Hopwood, evaluation itself is not a natural process; it tends to focus on the here and now and how to meet immediate goals, rather than assessing the long-term impact. He notes that the accountability side receives more attention when it comes to reporting than the learning side. This does not mean people are not learning, but rather that they do not always document the most important things that they learn. The problem lies in the lack of a systematic structure in place to assure lessons are shared; instead these lessons are "hidden away." This tendency of undocumented learning came up frequently in interviews: learning tends to be "implicit rather than explicit;" reporting on learning is "not systematic;" people "do not document learning, so we can't capitalize on it." These comments raise questions about learning especially when considering a project's sustainability or scaling-up a project.

Similar to Hopwood's logic, Emilie Aberlen adds that even if there is an evaluation team dedicated to evaluating all the projects of the AFD, reading a report takes time and teams all over the world often do not take time to read it. Patrice Dufour also notes that often those involved in selecting new projects read evaluations solely because their position requires it. Yet, they read only to check off a box on their to-do list; they rarely read reports with a sincere desire to learn.

Sue Soal stresses that the evaluation report should belong neither to the evaluator nor to the donor, but to the evaluated organization. This is the best way to promote learning within the organization. Furthermore, if the organization has ownership of the evaluation, they can adapt it in a way that allows them to improve their practices.

4. Dissemination of information

Traditionally, evaluation results have not always been viewed as information that is valuable to share. Even the shared results, are mostly contained to a handful of specialists, donors and evaluators. Luis Soberón, who works with the ELAC Project, which aims to promote inclusive evaluation practices in Latin America, spoke about a fairly new tendency to disseminate results to beneficiaries as well. Another key aspect about the information in evaluation results is the fashion in which it is presented. Soberón pointed at the language used in current evaluations. The highly technical language has presented challenges in the reception of this information; without a sufficient background in evaluation terminology, recipients may find reports difficult to understand. Intra-sectoral communication has also been neglected, Soberón continued, causing a break in the flow of information that could benefit evaluation. Similarly, Soberón finds that even within the same region there is no real homogeneity in the language used to publish results. Terminology, he adds, is fairly different amongst the countries of Latin America in which he has worked. This was present during the desk research; which is an indicative of a much wider concern. However, it is interesting to underline that OECD tried to solve this kind of issue in evaluation worldwide by publishing *Glossary of Key Terms in Evaluation and Results Based Management* in English, French and Spanish. This *Glossary* has not been mentioned directly by our interviewees; however, it seemed important to mention it here. While OECD *Glossary* might contribute to reducing misunderstandings related to different use of terms it can also make evaluation even more rigid and formalized which leaves less space for more innovative evaluation solutions adapted to a certain context or the rise of new concepts to name those solutions.

Pablo Rodriguez-Bilella points out that evaluations are not as transparent as they should be. Occasionally projects would cite confidentiality clauses in order to not to release key findings. This, Rodriguez-Bilella argues, hurts evaluation and the dissemination of information. He continues to say that if results are not positive, many evaluations are simply stored away, jeopardizing the learning process further. Again, we saw this phenomenon during the preliminary research; where transparency was a pivotal issue.

Sergio Martinic also highlights the role of an evaluator. He mentions that up to date, "the role of an evaluator remains unclear after she or he has finished a project". He questions whether evaluators could encourage better use and dissemination of results. Then adds that "there is the issue of resources, and all of these factors escape the capabilities of an evaluator". This he argues:

“an evaluator should inspire certain changes, but with the assumption that there are other dimensions, other roles that remain out of control for him or her.”

Sue Soal remarks that most of the dissemination does not occur because evaluation results remain in the hands of a few; i.e. donors, or implementing organizations. Indeed, the key link between evaluation and learning is how the information is disseminated. Ian Hopwood suggests that ‘the field of evaluation continues to skew towards accountability rather than learning”. This is a factor that has likely heavily conditioned the dissemination of information.

Almost everyone agreed that there is a great need to change how we disseminate evaluation. The evaluation sector could learn from regions like Latin America, which has started to make evaluations more transparent by utilizing online portals to publish results. These portals are government run and linked to public projects. Jan Van Ongevalle, however, underlines that these results should be used with caution and the conclusions to be taken with a certain level of awareness taking into consideration a particular context of each project. The question then becomes an issue of criteria: what then should be made public? Findings on the learning part? It then becomes unclear whose role it is to delineate these framework.

Evaluators expressed rather mixed opinions on evaluation networks. However, interviewees generally agree that there are venues and channels that keep evaluators in communication. Sue Saul, for example, indicates that monthly and yearly publications allow evaluators to stay up to date with new tendencies in their area; they also provide a space to reach out to other evaluators and share their own experiences. Mary Hall mentions the professionalism of international evaluation networks, which hints at the cited issue of closed communities mentioned by another interviewee. Pablo Rodriguez-Bilella suggests that evaluation networks require a wider audience if they want to be successful. He offers an example of always having the same groups of people at their organized meetings: “it feels like we are simply talking to each other and not outside our circle.” Oumoul Kharyi Ba Tall’s offered the most optimistic remarks on evaluation networks. She highlights the networks in his regions as continuously evolving and offering much needed resources to evaluators. She makes a special mention of EvalPartners, an international network that has offered pioneering services like online courses and resources for evaluators that transcend regions and language barriers. This is similar in Europe, where AFD participates actively with other organizations such as F3E and OECD eval-net are more prevalent. Evaluation networks, she continues, are often used as a tool to mobilize the community on focus areas and generate more traction because evaluators are encouraged to use their home networks. Ian Hopwood also emphasizes the link of networks and evaluation, which “enables one to answer difficult questions and improve results...in a world where people are increasingly nervous about admitting when things don’t work.” He also believes these networks provide evaluators with guidelines and encourage a much better understanding of the field of evaluation.

5. Soft Skills

Many of our interviewees stressed the importance of soft skills while conducting evaluations. Regardless of how the goals of the evaluation are set, the quality of the evaluation remains largely dependent on the evaluators' ability to observe a situation, communicate in a multicultural context, and reach conclusions.

One of our interviewees explains that the quality of the evaluator is at the heart of the evaluation because evaluation is a judgmental process and not a mechanical one. Evaluation involves an educated judgment. On the other hand, donors or organizations need to provide evaluators with reliable data so that they can do their job correctly. While data is essential, evaluators also rely upon their own ability to perceive a situation. For example, Patrice Dufour explains that evaluators do not always have the data they need at their disposal when they are assessing a project. An evaluator may be asked to see how a project has affected incomes in a given village. If there are no household income surveys available, a good evaluator learns to look for proxy indicators, such as the number of bicycles in a village or how many people have corrugated iron roofs instead of thatch roofs. Dufour concludes that the best evaluation practices often depend on small details, such as the flair that some people have for being able to read a situation or the savviness that comes with many years of experience. Of course a question then arises as to how to best administer proof that an evaluator's ability to perceive a situation is in fact legitimate grounds for issuing conclusions in the evaluation report. It is important to find a balance between the objectivity of hard data and the subjectivity of soft skills. The evaluation field could benefit from a deeper understanding of action-research, or the reflective and collaborative process that allows individuals to use their shared experiences to solve problems.

Many evaluators also addressed the need for cultural understanding when assessing projects. Luis Soberón explains that when evaluations are too heavily focused on technical requirement, they might lack a deep understanding of beneficiaries. In order to complete a successful evaluation, Soberón advocates learning about cultural context of the project. He shared an example of his work evaluating a food security program in an Andean community. In preparation for the project, he dedicated time to learning about the culture and society of the community where he would be working. Soberón argues that the creation of empathy is pivotal for his performance as an evaluator. These perspectives air similar thoughts to those of Totté's 4th generation of evaluation, which seeks to create a communal understanding between the local customs and norms and the implementing country's

Ian Hopwood echoes some of these same sentiments regarding the importance of understanding the local context. He explains that as an evaluator, it is important not to underestimate the ingenuity of people on the ground. The formal evaluation report may not always capture local innovation, but important learning can take place at a grassroots level. Therefore, the manner in which the evaluator engages with the local community has a tremendous impact on the evaluation outcomes. Working with local actors is also an essential part of Sue Soal's work at the CDRA. For her, evaluation is not a tool to impose judgment, but rather a tool of empowerment that can be adapted to the needs of the organization.

It is possible to change the norms or the goals of evaluation, but in the end, the quality of the evaluation will depend largely on the characteristics of the evaluator. Soft skills such as communication and cultural understanding are difficult to measure, but they may play a crucial role in building better evaluations.

Issue three: Innovation in Evaluation

6. Changing culture of evaluations

Many of our interviewees noted that evaluation culture is changing, even if it is changing at a relatively slow pace. And it is essential to keep in mind that the status quo of evaluation (as a donor driven process framed by the CAD principles) is still very strong. Although it is unlikely that evaluation will do away with the CAD or the logical framework, if not for it being an asset in structuring an evaluation as much as an engrained and common-way of approaching evaluation. Innovative ways of gaining a more nuanced perspective of evaluation should act as additional integral parts of the evaluation process as mentioned by Marc Totté. Emilie Aberlen, an evaluator at the AFD, for example notes that while the norm of evaluation is far from focusing on social change, this is the direction in which evaluation is moving and will increasingly become a focus. Furthermore, she highlighted that today's students of evaluation will likely focus on different aspects than today's evaluation practices.

Moreover, as a donor driven process, Florent Bedecarrats described how the perspective of evaluation has been closely linked to the focus and development intentions of the large financial institutions. Initially, evaluation was very technical and macro; it sought to ensure accountability in donor institutions' investments and in public procurement. In the 1980's and 1990's when the focus switch to Structural Adjustment Programs, evaluation took an approach that focused heavily on the price: was it just? were the methods used efficient? The five DAC criteria (relevance, efficiency, effectiveness, impact and sustainability), defined by the OECD, frame evaluation as a technical process constrained by the funder, while also being fairly participative. In 2005, the Paris Declaration on Aid Effectiveness further defined evaluation opening it up to become a more harmonious process. This suggests a shift toward a future of evaluation methodology that is more participatory.

A number of interviewees mentioned new trends in evaluation: slight switch in focus from output to outcome; more participation on the part of all stakeholders (especially the beneficiaries and implementing organizations). Moreover, evaluators observe an increased variety in the methods and tools of evaluation, notably an increased use of technology for data collection.

The advent of communication technologies such as Skype via the Internet has transformed evaluation. It has created the possibility to include a wider number of stakeholders and to facilitate a comparison amongst them. Florent Bedecarrats and Emilie Aberlen mention technological innovations used by the AFD in the evaluations. The AFD employs a lot of "geek" evaluation methods, notably by collecting on-the-ground data using computer-assisted interviews. In addition, the AFD increasingly uses satellite imagery in their evaluations. They have also started using Facebook as a means to reach out to stakeholders in various evaluations. The AFD has also used film and videos to broaden the perspective; Aberlen, however, nuanced that there is a limit to these methods replacing traditional methods of evaluation entirely because you still must be accountable for the millions of dollars you receive from a government or bi/multilateral institution. This need to ensure accountability could be a constraint on innovation. Nonetheless, they could be an important addition to the evaluation process to give a more holistic and all-encompassing perspective of the evaluation process.

An interesting aspect raised by Ian Hopwood relates to the information that is systematically left out of an evaluation report. He described his experience working on evaluation projects in Africa where the local actors on the ground were constantly learning and adopting practices as they learned them throughout the project cycle. As an evaluator, you help people learn about the issues at hand and how to adapt local methodology to the situation on the ground. Many times, however, the ingenuity of the local population is overlooked; but this is something that should not be underestimated. In fact, many formal evaluation reports do not capture local innovation. There is often a misperception that local capacities need to be built from the ground up, when in fact a great deal of local talent already exists. The question, therefore, is not how does one build local capacity, but rather, how does one further develop the talent that already exists. Reporting on the lessons learned locally could help when scaling up projects by formalizing what has worked and what has not.

7. Evaluation as a tool for social change

A central aspect of our work consisted of understanding to what extent evaluators believe that evaluation can be used as a tool for social change. Many interviewees are convinced that evaluation can and should be used as a tool for social change even if today it is often not the case. Each of them underlines different conditions that need to be met to achieve this goal (some of the following aspects evoke already mentioned ideas but relate them more directly to social change).

According to Jan Van Ongevalle, evaluation could become a very empowering tool but that depends on the way it is used. He believes that if evaluation asks the right questions (the ones that concern the needs of the communities and the way to empower them) and if it is well built into a project, notably from the beginning (aforementioned in the discussion on monitoring), it could be a tool for social change. Another evaluator underlines that for evaluation to promote social change, it requires disseminating information to the right stakeholders (most importantly the beneficiaries). Sergio Martinic, also believed that it is crucial to disseminate the results of the evaluation at a much wider range. Hopwood underlines that the ethics and standards of evaluation widely accepted by evaluators state that evaluation is important because it gives a voice to everyone, "not just stakeholders in the traditional sense." This could be used to induce social change. James Taylor emphasizes the importance of taking learning seriously if we want to use evaluation as a tool for social transformation (which according to him goes even further than social change because it creates a new system functioning according to new principles). Sue Soal believes that both external and internal evaluation can be used as a tool for social change. External evaluation is important because it helps to know what impacts the activities of the projects really have. Luis Soberón underlined another interesting aspect. Even though he thinks that evaluation can be used as a tool for social change, he believes that at least in Latin America the main obstacle for that to happen is related to governments. The potential of evaluation to induce social change might be hindered by an inadequate government framework. More precisely it means that lessons learnt from evaluation will not necessarily be translated effectively into improved policies or practices because of institutional shortcomings. This is an interesting aspect to consider. Marc Totté underlines that when using the techniques of the 4th generation of evaluation, social change is an integral part of the evaluation process, because the evaluation seeks to understand the perception that society has of the project's outcome, which axiomatically suggests a change in society's comportment or beliefs.

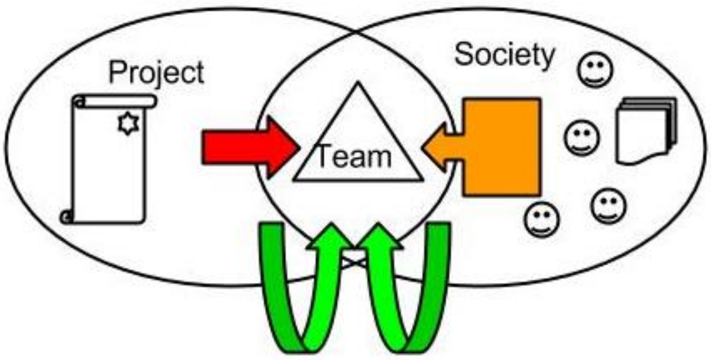


Figure 2: Marc Totté's 4th generation of evaluation

The scroll on the project side represents the pre-determined understanding of the project and its expected outcomes: the results that align with the logical framework and the donors' expectations of a projects use of funds. On society's side, however, norms are not only given by a charter or list of goals to accomplish but are often multiple and vary in sources from official documents to the opinions citizens hold. Each respective side brings their own value sets, which frequently differ from one another (red and orange arrows). The team conducting the project is where the two perspectives collide. Rather than the red arrow overpowering the conduct of a project based on the implementing countries understanding of societal norms (a frequent occurrence in today's evaluation methods), an approach of mutual understanding ought to be taken (the green arrows). The 4th generation described by Totté is an innovative method to look at a project through a common lens and with common goals and understanding defined by both the implementing country and the beneficiaries.

So, it can be seen that if certain conditions are met, evaluation could be used as a tool for social change. Moreover, for evaluation to become a powerful tool for social change it will be necessary to adapt new innovative tools and techniques that would be suitable for this goal.

Notwithstanding, it is essential to underline that interviewees highlight the difficulty of defining what social change is and how to measure it. Furthermore, it is necessary to clearly understand the difference between individual, organizational and social changes, which represent different realities.

It is important to note, however, a more skeptical opinion about evaluation being a tool for social change. For example, Mary Hall describes the other side of the coin, bringing up very relevant aspects to complement our analysis. Mary Hall underlines that we need to ask in the first place "should evaluation be used as a tool for social change". According to her, evaluation needs to have a mandate for social change because otherwise it could become manipulative. If we push social change on evaluation it might become less objective. Mary Hall explains that the aim of the project can include social change and evaluation can see if it has succeeded - but evaluation itself should not determine the nature of the change. Furthermore, if projects decide to include social change in their objectives, it should be clearly defined from the beginning of the project. Moreover, as achieving social change is a gradual process the time scale of a typical project would make it difficult for evaluation to capture the change unless it was undertaken well after the project ended. Even projects experience difficulties in producing social change, as it is a very long process; for evaluation, it would be even more difficult. As aforementioned, Mary Hall underlines that if evaluation were intended to produce a social change itself, ex-post external evaluation is unlikely to achieve this because at the end of the project it is too late for recommendations – although it could provide lessons learned for the future. Jan Van Ongevalle also mentions this idea. Oumoul Kharyi Ba Tall agrees that we should rather speak of social betterment and of evaluation as leading to improve public policies and in the end citizens' life but not directly social change.

Recommendations

8. Need to include beneficiaries

A substantial focus was placed during interviews on the need to better include beneficiaries in the evaluation process. In fact, the idea that beneficiaries should be a part of the evaluation process came up in nearly every interview.

Even though over recent years evaluation has attempted to become more inclusive and participatory while more actively involving the beneficiaries, there is still a long way to go in this direction. It is important to underline that here we use the definition of the beneficiaries that was indicated in the introduction of our work.

Many interviewees suggest that beneficiaries ought to be a central part of the evaluation, rather than just one component among many. Oumoul Khayri Ba Tall discussed the need for a "democratization" of the evaluation process by involving a greater range of stakeholders, notably the communities and citizens at large who have a stake in the evaluation and are affected by a project or program. As "consumers" of a project, Sue Soal of the CRDA believes that beneficiaries have the right to be involved in the evaluation. Furthermore, a focus on beneficiaries renders the line of demarcation between external and internal evaluation less distinct.

If the goal of the evaluation is to understand how a society has changed due to the project, this can only be done by discussing with the local community to discover how the project has changed their life and alleviated issues they used to face on a daily or regular basis. Marc Totté emphasized this concept by describing the evaluation process at Inter-mondes. The first step is to ask the local population: "What are your sufferings and your greatest worries? What is the nightmare that wakes you up at night or keeps you from sleeping?" The responses give you a sense of the underlying problems and issues in a society. The next step asks if the project alleviated these sufferings and worries or if it responded to the issues and problems mentioned. Simply asking: *did the project help you?* will only give a "yes" or "no" answer; it gives no deeper understanding about the changes taking place in the society and the underlying progress or development.

Having a more comprehensive perspective of a project is essential; the beneficiaries give an understanding of the true impact of a project beyond what can be measured by quantitative data. It measures to what extent there has been a change in comportment and beliefs due to the project. In the words of Sue Soal: "good development work is transitioning from being a transfer of resources to being a transformation of relationship." Nonetheless, this perspective implies an active will to understand the beliefs and cultural aspects of a community and how they may differ from the project team or funder's worldviews. The fourth generation of evaluation proposed by Marc Totté approaches the project from one comprehensive perspective that takes into account not only the pre-defined goals of the funding organization and the strictly defined logical framework, but also the cultural subtleties and norms of the local community. This cultural aspect is very important.

Ian Hopwood is also committed to involving the opinions and perspectives of local actors in his evaluations. Particularly, these local actors increasingly include youth. For example, in May 2015, Hopwood will participate in an evaluation conference with four young West Africans. The youth will have the opportunity to present their ideas and contribute to conference participants learning on evaluation. Hopwood cites Kim Sabo Flores' endeavor to involve youth in evaluation as well. She is using the evaluation to promote learning in a in a youth-friendly manner through theater and other participatory methods. Drama productions allow youth to express themselves and their views on the issues that affect their lives. Other institutions such as the American Evaluation Association are also working on involving youth in evaluations. The American Evaluation Association has created a youth sub-group to work on evaluation issues. Organizations, such as FHI 360 (a U.S. based NGO), have adopted a youth-involved approach to evaluation as well with their Life-Skills program.

John Healy highlights the nuance between the approaches that institutions take versus NGOs when it comes to the inclusion of beneficiaries in the evaluation process. While institutions such as the AFD admit that they focus more on the government or institutions in a partner country as the beneficiaries, organizations, especially those involved in social movements or advocacy movements, are much more likely to involve beneficiaries as a central stakeholder when conducting or commissioning evaluations.

Given the emphasis placed on the participation of beneficiaries by our interviewees and the will to utilize evaluation as a tool for social change, a greater inclusion of the project or program community ought to be a central aspect of evaluation. This inclusion ought not to be merely secondary or an afterthought to other qualitative and quantitative on-the-ground research as part of the evaluation process, but a priority. To understand the true impact and sustainability of a program, it's necessary to understand the influence the project or program had on the daily lives of the local community and how their comportments and beliefs have changed in conjunction with or in reaction to the project and program.

9. Need to focus on the process

Across the board, experts emphasized the need to shift the focus on the process, rather than to simply on the results. The issue arose with reference to the logical framework, which for some experts might deviate attention away from the means. Mark Totté points out that the objective of evaluation weights on seeing the conclusions. He argues that evaluation should take a more active interest in understanding the movement between states, the process of change rather than whether or not the elements of the logical framework can be deemed completed. " There is a tendency to follow exactly a logical framework, to think in a very structured way about evaluation." Jan Van Ongenvalle illustrate this point arguing that the logical framework is almost an imposition, and there is a diffusion of responsibility as to who are the appropriate actors that need to adjust or redefine the impacts in an ongoing project. The logical framework is certainly a useful tool to construct intervention, however, it must be accompanied with a wider understanding of crucial aspects or tools that might not fall neatly into a checkbox category; and move parallel to the conception and process of a particular project.

Occidental countries tend to be more focused on evaluating to understand the results of a particular project. Because these countries are also generally the donors in what we have defined as a "donor driven process." This propagates this particular focus of evaluation. Yet, this perspective leaves out the crucial aspect that change is a movement from one state of being to another. It necessitates a movement from state "A" to state "B." The end product of a "results focused" evaluation, however, gravely overlooks this movement between point "A" and point "B." The interviewees would like to see a deeper exploration of the transit route between the points. Comparing a starting and ending point is not sufficient because it does not show the entire picture. While experts do not have a consensus on how to incorporate this new approach, they do suggest a steady move towards more qualitative evaluations. There is also a call for more involved, and locally driven evaluations that can be more representatives of their respective situations.

IV. Interview Assessment: Regional Differences

General Overview

Another aim of our work was to look at how evaluation methods, approaches, objectives and challenges differ according to regions. Ideally, we wanted to deduce these differences from our interviewees' responses without asking them if they believe a regional or linguistic difference exists. However, we eventually decided to introduce this question explicitly in our interview matrix as interviewees did not cover this subject in their discourse and it was not possible to deduce regional differences when analyzing their responses. This suggests that regional differences may be an afterthought to evaluators. They have a set of predefined tools that they apply and adapt to various geographical situations.

After speaking with actors from various geographic and linguistic backgrounds, we found little evidence to suggest clear regional differences in the evaluation field. Even when asked directly, many interviewees affirmed that they observed little difference. Some relayed that while in theory a difference between regions exists, in reality it is difficult to observe it. Others described the international evaluation community as a brotherhood that uses a common terminology regardless of where they are working in the world. This idea of an "evaluation brotherhood" is quite controversial, as it suggests a set of homogenous goals that not all stakeholders in the evaluation process are likely to accept. Evaluators who work for large donors may be more likely to perceive evaluation as a brotherhood, because, as James Taylor of the Community Development Resource Association (CDRA) points out, evaluations are often done to satisfy the requirements of donors, which leads evaluations to be "fairly standardized according the principles of the powerful." Moreover, Mary Hall amongst others underlines that the major donors use global systems and global templates. For example, in the UN the same templates are used in many different countries and this could explain why regional differences are less apparent.

Nevertheless, at a grassroots level, evaluators recognize cultural differences in how evaluations are perceived and interpreted. Patrice Dufour of AidImpact suggests that the evaluation process itself does not change from one region to the next, but how people respond to the evaluation is linked to culture. Moreover, Florent Bedecarrats seemed to think that the largest difference is chronological, which is in congruence with the idea his co-worker Emilie Aberlen presented on the future of evaluation becoming more participatory. Oumoul Kharyi Ba Tall underlines that although different countries might view evaluation differently, these differences might not be related to geography, but rather to the political characteristics of a specific country. For example, since democratic countries must prove their accountability to taxpayers, they may place a greater emphasis on evaluation. One should avoid making broad generalizations about a specific region or culture, since a single region often contains tremendous cultural diversity. However, even while keeping this diversity in mind, it is possible to look for regional trends.

Regional differences were difficult to capture; however, linguistic ones were even more difficult to deduce. There seemed to be little reason to divide Africa and Asia into English and French speaking regions, based on their colonial heritage. It is difficult to develop clear-cut between regional or linguistic norms in evaluation simply because of the multiplicity of actors and cultures at play in a given project. For example, if a British NGO is implementing a project in francophone West Africa, the evaluation may be influenced by Anglo-Saxon, French, and African traditions. However, some evidence suggests that linguistic differences do impact evaluations in Europe and North America. A presentation of these differences and other geographical aspects that we managed to deduce from our interviews will follow.

Regional Characteristics

1. Anglophone Europe – USA – Pacific

According to the literature review that was presented at the beginning of our work, this region zone was expected to have a more *accountability orientated* and *positivist* approach (where one can obtain objective knowledge through observation). Interviews did not really reveal these two characteristics (this, however, does not mean that they are absent in the reality). Nevertheless, some interviews underlined other interesting aspects that we would like to highlight.

Mary Hall notes a certain difference between the Anglo-Saxon approach and the French approach to evaluation. She believes these differences between Anglo-Saxon and French evaluation approaches transcend into greater linguistic differences between territories. For example, Mary Hall believes that Anglo-Saxon tradition was very present while creating the monitoring system and this is why monitoring remains influenced by this tradition. She considers the Anglo-Saxon approach to be more pragmatic and more results oriented, albeit these results can be social change or whatever was agreed as the purpose of the projects. Another interviewee underlined that Anglo-Saxon approach is considered to be more factual.

Jan Van Ongevalle and John Healy both mentioned that in the US randomized control trials are often considered as a golden standard. However, Jan Van Ongevalle clearly states that he does not agree with the fact that considering randomized control trials as a golden standard is a positive aspect that can enable learning. He actually finds this hierarchy assigned to evaluation methods a setback. The evaluation question and context should determine the most suitable method and not the other way around. Some interviewees mention that in the US, evaluation is more a profession, which might result in a willingness to make it a "club"; while our interviewee saw this as a positive aspect to enable sharing, our analysis shows that this could have negative repercussions and actually inhibit learning and sharing by creating exclusivity.

2. Non-Anglophone Europe

It was difficult to clearly identify the characteristics of this region. We consider it relevant to also address other European countries that cannot be qualified as English or French speaking (here we exclude the United Kingdom that belongs to a previous group). Thus we have expanded the region to include other European countries given commentaries made by our interviewees.

A. France

While a particular "French" culture of evaluation did not appear in our interviews, some comments were mentioned. For example, one interviewee underlined that French evaluation culture in theory is considered to be more analytical.

B. Wider Europe

According to Emilie Aberlen, Northern European countries are very keen on publishing and discussing project failures in evaluations as a means to learning and improving their process of evaluation. Moreover, John Healy perceives evaluation in Europe to be less well defined and a smaller community. Evaluation is done through the state bureaucracy process; as such, the results of evaluation tend to be more used to back up decisions that are already made. However, another interviewee underlines an important aspect: evaluation community in Europe is not homogeneous. For example, the evaluation tradition is strong in the North but was almost absent in the South of Europe for a certain period of time.

3. Asia and Africa

Despite the growing emphasis on the need for locally driven development solutions, the Western World continues to play a significant role in setting the development agenda of the global south. Our interviewees confirmed the need for local ownership of evaluation in developing countries. The following sections focus on our interviewees' perceptions of evaluation trends in Africa and Asia.

C. Africa

Our interviewees did not highlight any trends specific to Africa. This could be partially due to the fact that Africa is a diverse region without a homogenous culture. Another explanation for a lack of cultural specifics is the fact that, as James Taylor noted, outsiders have defined the evaluation process, not by local African institutions. The practice of evaluation in Africa, at least for the international organizations, remains tightly framed by the formulation of the terms of reference, and the practices and language of the donor organization. Hence, according to Patrice Dufour, a former World Bank evaluation officer, the approach of the World Bank consultants in Africa was much more technical and macro, focusing on questions of investment and procurement policies.

Evaluators who had worked in Africa shared some of their observations in the region. Patrice Dufour explains that in the Middle East and North Africa (MENA) it is important for individuals not to "lose face" before their colleagues; therefore, evaluations can be perceived in a negative light. There is a fine line between giving recommendations and unintentionally insulting someone.

Ian Hopwood also spoke of his experiences working on evaluations in West Africa. He argues that the ingenuity and creativity of local actors in West Africa is one of the continents greatest resources, but that ingenuity is sometimes overlooked in the formal evaluation process. In Hopwood's words, international organizations and donors need to learn that the "biggest resource in Africa is the time and commitment of individuals [Africans]." The flexibility and resilience that allows individuals to successfully run projects is something that evaluators may notice, but this implicit type of learning might not be included in the official evaluation report.

Interestingly, when expressing his views on the relationships between associations and donors, James Taylor uses a Western/Southern dichotomy that also fits the rich/poor lines of argumentation. This could lead us to think that the evaluation process in Africa is not only a technical issue but also a very sensitive political issue.

B. Asia

Mark Totté suggests that Asia has a more holistic approach to evaluation because their culture encourages it; they would be more inclined to look at the process, whereas occidental countries are much more orientated towards the results. The emphasis on learning rather than accountability is also related to cultural norms regarding politeness. Patrice Dufour explains that similarly to when working in the MENA region, it is important for evaluators to be aware of how their critiques will be interpreted culturally in Asia. What might be perceived as advice in a European context can seem like a personal attack in an Asian context. On the other hand, Dufour argues that it is important to strike a balance between respecting culture and providing an honest review of the problems that come to light during an evaluation. Cultural awareness should not come at the expense of an accurate evaluation.

Asia is a diverse continent with a variety of cultures, norms, and standards; thus a holistic approach to learning is not always the key focus in the region. For example, when working on projects funded by the Chinese government, Emilie Aberlen noted that there was greater emphasis on accountability than on learning. In this context, evaluators were more results oriented; they wanted to highlight the successes to the government in order to justify the aid. However, this results focused orientation may be less of a cultural reality than a political reality. Government workers may be interested in providing results to their superiors, not because of any cultural inclination, but because their job requires that they produce results.

4. Latin America

In Latin America the tendencies are far from being homogeneous. The culture of evaluation is fairly new and has evolved rather slowly. Evaluation has been tied to governmental projects as a mere reflex. In other words, there was no indication of an organic bloom of evaluation in the region. This, in itself, has presented challenges for the development of a culture of evaluation. Most expert evaluators agree that the focus should be now on shining light into the benefits of evaluation, beyond its mere ability to render a project accountable. The region, however, has been benefiting from organizations like EvalPartners. They have provided evaluators with resources, and allowing them to take courses online to improve their skills. Leading evaluators also understand that the region's heterogeneity poses particular challenges. Luis Soberón, a leading Professor of Evaluation at the Catholic University of Peru, has worked in more than six countries in the region. His remarks left no doubt that a successful evaluation culture is heavily linked with the degree of the country's democratic system. Latin America is fertile soil for evaluation, he concluded, but we must increase our dialogue with each society if we would like to see a spirit of change. Another interviewee also added that evaluation culture in Latin America is considered to be more empirical, meaning that there is a strong belief that what worked in one country cannot really be reproduced in another one. Moreover, Florent Bedecarrats, when asked about regional differences, stated that the popular education movement has made evaluation more participatory in the Latin American region. This fact, according to him, was the only regional difference he denoted. Overall, the region seems to be in a good balance, although interviewees to not take the progress achieved so far for granted. All of them agree that there is a long road into transforming the evaluation culture in Latin America

V. CONCLUSION

General conclusions

Overall, our literature review and interviews with evaluators working all over the world and in a variety of languages have pointed to a number of interesting characteristics regarding the norms and practices of evaluation. From this, we draw a series of conclusions.

Imposing a universal definition of evaluation in diverse cultural setting

Contrary to our initial mandate, it seems that regional and linguistic differences are less apparent when it comes to evaluation. This could be due to the harmonization of evaluation approaches under institutions, such as the UN, OECD and World Bank, defining evaluation norms and criteria. It could also be strongly affected by the fact that evaluation remains largely a donor driven process, and those donors tend to come from the global North. This standardization of evaluation, however, has its limits as the partner countries where evaluation is conducted have varying cultural and historical contexts that ought to be considered. This is a central point that many smaller evaluation organizations and NGOs, such as the CDRA, highlighted in their discussion of the evaluation process. These cultural differences can affect not only the process of evaluation (whether it is more analytic, pragmatic, utilitarian, holistic or a learning opportunity), but also the way it is perceived. While our analysis shows that regional differences in evaluation culture are much smaller than we expected at the beginning of our project, it would be inaccurate to advocate that five regions identified have identical evaluation cultures. This is most certainly not the case; however, we do not have enough information to identify specific evaluation cultures for various regions. The lack of clearly specified regional differences can also come from the limits of our analysis that will be presented at the very end of the work.

Furthermore, the need for a cultural understanding of the project community and country is interconnected with the issue raised of needing to include beneficiaries as a more integral part of the evaluation process. Many interviewees noted that while this would be desirable, it is often not done because there is neither enough time nor money to complete this step. Including beneficiaries as a true stakeholder in the process, would give a more complete understanding to the impact and sustainability of the project. Furthermore, beneficiaries could provide a unique perspective of the realization of a project; if this vantage point were more often included in formal reports and effectively taken into consideration, it could provide an additional opportunity for learning.

The type of organization determines the definition

Although we did not find clear regional differences when asking evaluators how they defined evaluation, we did see variations based on the kind of organization in which they worked. Evaluators who had spent the majority of their time working for large international or national donors, such as the EU, World Bank or AFD, highlighted the formal aspects of evaluation. They tended to emphasize the fact that evaluations are a donor driven process strongly influenced by OECD-DAC criteria. Evaluators working for smaller organizations, on the other hand, were much more likely to highlight the role that learning plays in an evaluation. While many evaluators working for smaller non-governmental organizations still agreed that donors largely influence the formal evaluation process, they also stressed the crucial nature of the informal learning process to promote impactful development work.

The institutional background of evaluators seems to heavily influence the manner in which they speak about their work. Therefore, rather than seeking out regional differences, it would be interesting to delve further into the issue of institutional differences. Two different evaluators from the same region could have completely different perspectives on the evaluation process depending on whether they work primarily on evaluations funded by large international donors, such as the World Bank, or if they work primarily on evaluations funded by smaller organizations or as an independent contractor. If we were to take this research a step further, we would ask more questions about the influence of institutional backgrounds, rather than simply regional differences.

The need to expand the focus beyond evaluators

The fact that evaluators are strongly influenced by the type of organization or donor with which they typically work, supports our initial argument regarding the need to expand this research to include additional stakeholders. Although focusing on evaluators was logical in producing a preliminary benchmark on evaluation, it limited our ability to gain a holistic understanding of evaluation on a global scale. Moreover, expanding the analysis to other stakeholders could also permit to see more clearly power dynamics and interactions that exist among different stakeholders of evaluation. This is a crucial aspect to develop in the future research work.

The fact that our research focused on evaluators, who typically work on a variety of projects around the globe, may have contributed to the lack of evidence supporting regional differences. If we had focused on organizations implementing projects or on the project beneficiaries, we may have found clear-cut regional differences. The majority of evaluators we spoke with stressed the importance of culture, which suggests that organizations with different cultural backgrounds may express different feelings and perceptions regarding the role of evaluation. Again, however, the size and nature of the organization is likely to influence its views on this aspect of evaluation. For example, a small grassroots NGO likely understands evaluation differently than a large multi-national organization.

While it is interesting to consider how evaluation differs on a global scale, one must be careful to avoid sweeping generalizations. Each region we considered contains a rich cultural and linguistic diversity. Even if we do not find clear evaluation trends at the regional level, there may be distinct evaluation cultures at the country or even community level. By expanding the focus beyond evaluators, we would be able to gain a more comprehensive understanding of evaluation. Broadening the scope of the research would also make it easier to identify subtle nuances in how evaluation is used and perceived.

The need for a more comprehensive and inclusive approach to evaluation

In addition to including more stakeholders in the analysis of evaluation, it is also important to consider the variety of ways in which evaluation is used. Although we initially focused on ex-post external evaluation, many of our interviewees suggested the need to consider and implement monitoring and evaluation since the conception of a project. Monitoring and evaluation go hand-in-hand; therefore, the relationship should not be an afterthought towards the end of a project, but rather built into the entire framework of a project and considered throughout the project cycle. Furthermore, monitoring and evaluation, when used in this way, can be an instrumental tool for social change.

Other interviewees pointed out that the goals of empowerment and learning may be better accomplished through internal evaluation. When implementing organizations own the evaluation process, they can use it as a tool for social change. This assumes, however, that all organizations display willingness to self-evaluate and to learn from their mistakes, when in fact, this might not always be the case. Therefore, we must return to the dual nature of evaluation: accountability and learning. Evaluations should promote organizational learning, but they must also hold people accountable at the project level. Better understanding how these two disparate sides of evaluation fit together is essential for transforming evaluation into a tool for social change.

Requisites for evaluation as a tool for social change

Although the majority of our interviewees believe that evaluation can be used as a tool of social change, they also highlight that it is not perceived and used as such in a majority of cases. Interviewees offered a variety of suggestions on how evaluation can achieve this new goal. Evaluation as a tool for social change requires: asking the right questions centered on the needs of the communities and how to empower local populations; an increased dissemination of information to stakeholders (where beneficiaries are no longer considered as second rate stakeholders); a more rigorous and integrated effort to understand the local community and the evolution of their customs and comportments as part of evaluation; and an increased emphasis on learning. Nonetheless, our interviewees highlight not only the difficulty to define what one means by social change, but also how to measure it. In fact, interviewees gave a wide spectrum of opinions on how they interpret the term "social change." Moreover, some interviewees offered skeptical opinions, also analyzed in our work, which provide a starting point to consider different facets of the issue.

An interesting and important aspect of evaluation mentioned by interviewees was the role of soft skills for evaluators. The skill-set and knowledge brought by the individual evaluators to the process often determines the quality of an evaluation, regardless of donor driven demands or principles. This is largely due to the judgmental nature of the process; the manner in which evaluators interpret data will strongly influence their report. Thus, more on-the-ground experience in evaluation is a clear asset as evaluators are constantly learning-by-doing. Furthermore, having a cultural understanding will increase the quality of an evaluation as it enables an empathetic understanding of projects.

A lack of learning consistently came up as a critique in our interviews. It is not only seen as an essential aspect of evaluation, but also as a means for evaluation to enact social change. The need to better promote learning was a common theme throughout the interview process. For learning through evaluation to enact social change, evaluation reports must take into consideration the learning experiences of the local population. This process of only individual evaluators learning is insufficient. If the evaluator is an outside consultant, when they leave all the knowledge leaves with them. It is imperative that this knowledge and the learning experiences are shared with implementing organization. Organizations could learn from these experiences to know what to incorporate or to avoid for future projects and evaluations. While those who work in evaluation have a sharing profile by nature, interviewees consistently mentioned that too many evaluations get sent out, placed on a shelf, and collect dust. There needs to be more communication across different stakeholders so as to encourage a more active sharing of information. There ought to be a systematic shift from individual learning that very often concerns evaluators to organizational learning in implementing organizations. Organizational learning requires thinking, strategy, and management incentive; fundamentally, this requires radical decisions from the top-levels.

Caveats to research

With a limited time and no budget, our research was albeit limited. Had we had more time, we could have widened the scope of our project to include perspectives of a wider number of actors. As a thought experiment, we created a broad interview matrix (see annex), which includes possible interview questions for donors and organizations that implement projects on our same three themes on which we consulted evaluators. We believe that these two additional actors play an integral role in the evaluation process and shape its methodology and focus.

An additional caveat to our research lies in the issues we confronted with our access to evaluators in various regions, largely due to non-responses. For example for Latin America, while we were able to secure three interviews (Argentina, Chile and Peru), these evaluators' experience did not cover Brazil. Not only is Brazil the largest country in Latin America, but also an important actor in development and social change in the region. Furthermore, Mexico was not included in our report because we were unable to reach a contact from the country.

Similarly for the francophone region, we received no response from our contacts in Quebec; so while it was initially included in our report with some very interesting perspectives from our literature review, we were not able to attain a direct opinion. This may have skewed our analysis into thinking that evaluation cultures are more homogenous across occidental francophone countries in reality than they originally appeared in our literature review.

A weakness of our Asia assessment is that we did not have the opportunity to interview any evaluators from Asia, only European evaluators who had worked in Asia. This may give a biased perspective of Asian evaluation culture, influenced by European norms. This was also the case for our analysis of evaluation in Africa. While we were able to speak with evaluators from South Africa and Mauritania, our only perspectives for East Africa came from Europeans who had worked in these regions. The Europe-US-Pacific region analysis lacks direct perspective due to this same issue. We were unable to contact an evaluator from the Pacific region. While we were able to interview an evaluator who had worked in the US for a significant period of time, we did not have access to someone currently working in evaluation in the US.

The lack of direct perspective for some of our regions limits the extent to which our analysis can be extrapolated as an accurate global view of evaluation today. Further research ought to be done to fill the caveats of our research with on-the-ground experts that come from the local communities. Additional interviews should include a wide range of stakeholders to gain a better understanding of how evaluations are interpreted and received in the various regions. This would give a more comprehensive and accurate understanding of differing global norms, standards, and focal points of evaluation today. Nonetheless, the comments and considerations we did receive cover a wide variety of geographical areas and suggestions such as increased sharing and learning and a heavier focus on beneficiaries ought to be taken into consideration across the globe for a more meaningful evaluation process.

VI. REFERENCES

AFD. (2014). L'évaluation en évolution: Pratiques et enjeux de l'évaluation dans le contexte de l'éficacité du développement (Conférences et Séminaires). Paris: Agence Française du développement.

African Evaluation Guidelines - Standards and Norms. (2007, January 1). Retrieved March 10, 2015, from http://www.ader-evaluare.ro/docs/African Evaluation Association.pdf

Barefoot Collective. (2009). Barefoot Guide to Working with Organizations and Social Change. (2) p. 1-174

Brohman, J. (1995). Universalism, Eurocentrism, and Ideological Bias in Development Studies: From Modernisation to Neoliberalism. Third World Quarterly. 16(1). 121-140.

Ceori, D; Pozo Solis, A (2007). Institutionalizing Monitoring and Evaluation in IFAD-Co-financed Project in Latin America and the Caribbean. Summary of surveys of directors and officers of monitoring and evaluation units, 2004 and 2007. Retrieved March 10, 2015 from http://preval.org/files/resumen_ejecutivo_institucionalizacion[1].english_08.pdf

Christie, J., (2008, November). "L'évaluation des actions de développement – Pourquoi? De quoi? Pour qui? Rapport sur les défis de l'évaluation tels que perçus par certaines OSC canadiennes pour le Conseil canadien pour la coopération internationale. 1-29

Conlin, S., & Stirrat, R. (2008). Current Challenges in Development Evaluation.Evaluation, 14(2), 193-208.

Conseil scientifique de l'évaluation. (1996). Petit guide de l'évaluation des politiques publiques. Bulletin des bibliothèques de France [en ligne], n° 3, 1997 [consulté le 27 avril 2015]. Disponible sur le Web : http://bbf.enssib.fr/consulter/bbf-1997-03-0106-002

Delivering on commitments. (n.d.). Retrieved March 5, 2015, from http://www.mdgfund.org/content/monitoringandevaluation

Duran, P., & Monnier, É. (1992). Le développement de l'évaluation en France. Nécessités techniques et exigences politiques. Revue Française de Science Politique, 42(2), 235–262. http://doi.org/10.3406/rfsp.1992.404295

Earl, S., Smutylo, T., and Carden, F. (2001). Outcome Mapping; Building Learning and Reflection into Development Programs. Ottawa: IDRC.

Estrella, M. & Gaventa, J. (1998). Who Counts Reality? Participatory Monitoring and Evaluation: A Literature Review, IDS Working Paper 70, Brighton: IDS.

EvalPartners. (n.d.). 2015 declared as the International Year of Evaluation | MY M&E. Retrieved March 5, 2015, from http://mymande.org/evalyear/Declaring_2015_as_the_International_Year_of_Evaluation

F3E. (2014a, October). Les "approches orientées changement" : de quoi s'agit-il? Retrieved May 2, 2015, from http://f3e.asso.fr/media/transfer/doc/presentation_aoc_v3.pdf

F3E. (2014b, December 16). La Capitalisation des expériences. Retrieved May 2, 2015, from http://f3e.asso.fr/ressource/44

Gasper, D. (2000), Evaluating the 'logical framework approach' towards learning-oriented development evaluation. Public Admin. Dev., 20: 17–28.

La Documentation française. (2044, August). Chronologie [dossier d'actualité]. Retrieved March 10, 2015, from http://www.ladocumentationfrancaise.fr/dossiers/evaluation-politiques-publiques/chronologie.shtml#

Lopez Acevedo, G; Rivera, K; Lima, L; Hwang, H (2010). Fifth Conference of the Latin America and the Caribbean Monitoring and Evaluation (M&E) Network Challenges in Monitoring and Evaluation: An Opportunity to Institutionalize M&E Systems. Retrieved March 10, 2015 from http://siteresources.worldbank.org/INTLACREGTOPPOVANA/Resources/840442-1255045653465/Challenges_in_M&E_Book.pdf

Modernizing Foreign Assistance Network (MFNA). (2014). The Way Forward: A Reform Agenda for 2014 and Beyond. Retrieved March 10, 2015, from http://www.modernizeaid.net/documents/MFAN_Policy_Paper_April_2014.pdf

OECD, "The Paris Declaration on Aid Effectiveness". (2005, March 2). Available at http://www.oecd.org/dac/effectiveness/34428351.pdf

OECD. (2010). Belgium (2010) DAC Peer Review - Main Findings and Recommendations. Retrieved April 12, 2015, from http://www.oecd.org/dac/peer-reviews/belgium2010dacpeerreview-mainfindingsandrecommendations.htm

Participatory Research in Asia (PRIA). International Perspectives in Participatory Monitoring and Evaluation. (2015, January 1). Retrieved April 11, 2015, from http://www.pria.org/international_perspectives_monitoring.htm

Piccio, L. (2014, September 4). How much do impact evaluations cost? Retrieved May 2, 2015, from https://www.devex.com/news/how-much-do-impact-evaluations-cost-84267

Picciotto, R. (2007). The New Environment For Development Evaluation. American Journal of Evaluation, 509-521.

Rogers, K. (2015, March 3). M&E: Moving toward a culture of learning? Retrieved March 5, 2015, from https://www.devex.com/news/m-e-moving-toward-a-culture-of-learning-85612

Shore, K. J. (2008). La mesure du changement social. Retrieved March 10, 2015, from http://www.idrc.ca/FR/Programs/Evaluation/Pages/ArticleDetails.aspx?PublicationID=745

Société Française de l'Evaluation. (2013). Evaluer une action publique. Paris: Société Française de l'Evaluation. Retrieved from http://www.sfe-asso.fr/intranet/ckfinder/userfiles/files/Bibliotheque/SFE_Guide_Elus.pdf

Storm, S. (2013) Fomento de Capacidades en Evaluación en países seleccionados de Latinoamérica (FOCEVAL). Retrieved March 10, 2015 from https://www.giz.de/en/downloads/giz2013-sp-evaluacion-foceval-latinoamerica.pdf

Struder, N. (Ed.). (2010). Évaluation des politiques publiques sanitaires et sociales. Revue Fraçaise Des Affaires Sociales, (1-2).

The American Evaluation Association (AEA) (2011). Public Statement on Cultural Competence in Evaluation. http://www.eval.org/p/cm/ld/fid=92

Thomas, V., and X. Luo. (2012). Multilateral Banks and the Development Process: Vital Links in the Results Chain. New Brunswick, NJ: Transaction Publishers.

Wang, Z. (2014). Evaluation and Results in Asia. In Evaluation for Better Results. Manila: The Asian Development Bank.

Zarinpoush, F., & Imagine Canada (Organisme). (2006). Guide d'évaluation de projet à l'intention des organismes sans but lucratif: méthodes et étapes fondamentales pour procéder à l'évaluation de projets. Toronto: Imagine Canada

VI. ANNEXES

Annex 1: Broad Interview Matrix

ISSUE 1: Defining Evaluation

Donors of development projects and programs

In a "door defined process", it is crucial to understand how donors define, understand, and use evaluation. By researching the role donors play in evaluation, one can better understand the power dynamics on which evaluation depends and how these power dynamics shape evaluation

- To what extent is evaluation a required criteria in projects and programs you fund? If not required, what inclines you to request evaluation reports?
- As a donor, in which roles of evaluation are you most interested (accountability, learning, another aspect)?
- What are the main reasons that encourage you to invest in evaluation?
- Who uses the reports of evaluation in your organization? How is it done?
- What evaluation criterion needs to be met in order for you fund a project?
- How do you choose your evaluators? What do you expect from evaluators?
- In your opinion, if and to what extent could evaluation be used as a tool for social change?

Evaluators

Evaluators work on the "front lines" of the evaluation process, employing a variety of methodologies to assess projects. In asking evaluators to define evaluation, we seek to identify common perceptions, themes, and goals that influence the evaluation process.

- What is evaluation for you?
- What is your main goal as an evaluator?
- How is this goal determined? (By donors, organizations that implement the project, or personally identified needs?)
- In your opinion, if and to what extent could evaluation be used as a tool for social change?
- To what extent do donors influence the manner in which you conduct an evaluation (methodology, goals, points included in evaluation reports)?
- How do project beneficiaries influence the evaluation you are conducting?
- To what extent do project beneficiaries participate in evaluations?
- In your work, have you noticed any regional differences in approaches to or methods of evaluation?
- To what extent do you feel organizations adopt recommendations?
- Do you have a post-evaluation follow-up with the organizations with which you work? If so, how does this work? If not, do you believe it would be useful?

Implementing Organizations

Organizations also play a central role in evaluation, since it is their work, which is being evaluated. It is important to understand how organizations feel about external, ex-post evaluation, and how this may impact while planning and monitoring the implementation of projects. Understanding how organizations identify the aims of evaluation will help to comprehensively define the concept of evaluation.

- What do you expect from evaluation?
- What role does evaluation play in your projects? What role could it play?
- How do you understand the aim of evaluation?
- What role do you play in an external evaluation? How are you involved in it? Do you complement it with another type of evaluation?
- In your opinion, if and to what extent could evaluation be used as a tool for individual and organizational social change?
- Do you have any personal experiences that illustrate how evaluation has impacted your organization's work?

ISSUE 2: Evaluation and learning

Donors of development projects and programs

To understand the relationship between evaluation and knowledge acquisition, one must establish if learning is a part of the evaluation process. Evaluations could be used to show donors not only how their money was spent, but also, possibly, the impact of their investment. Furthermore, donors may request evaluations primarily for tracking purposes, or they may use evaluations to identify best practices in the development field. Asking donors to identify what, if anything, they want to learn from evaluations will help clarify the balance between accountability and learning. It is possible that donors request evaluations only for the sake of holding organizations accountable for how their money is spent. On the contrary, evaluations could also be used to promote learning on the best kinds of projects to fund.

- What, if anything, do you learn through evaluation? Can you give specific examples of why learning may or may not be a goal of the evaluation?
- Who, in your opinion, should learn from evaluation?
- How do you use the results of evaluation?
- Does the knowledge acquired through evaluation lead to changes in where you donate your money? If so, how? If not, what are the barriers to changing funding norms based on evaluation outcomes?
- To what extent do you communicate with other stakeholders during and after the evaluation process? Do you encourage this kind of interaction? How does this effect learning?
- Does communication continue after the evaluation is complete? If so, with whom do you communicate? What is the objective of this communication?

Evaluators

Evaluators are uniquely situated to identify the links between evaluation and learning as they witness what occurs on the ground and compile a report that includes recommendations for the implementing organization. They will have an idea not only of who requests evaluations, but what purpose the evaluations serve both for organizations and donors. Therefore, evaluators can provide crucial insight into the extent to which learning is an aim of evaluation.

- What is being learned through evaluation?
- Who learns from evaluation? Who, in your opinion, should learn from evaluation?
- What do you do with the results of evaluation?

- Does the knowledge acquired through evaluation lead to changes in how evaluations are conducted (demands from donors or you own methodology or focus)? If so, how? If not, what you think are the barriers to changing evaluation practices?
- To what extent do you involve other stakeholders in evaluation procedure (donors, communities)? Do you think there ought to be a limit to the involvement of these other stakeholders in the evaluation process? How do you communicate with other stakeholders? How does this effect learning (your learning, donors' learning and the learning of the beneficiaries)?
- What role, if any, do beneficiaries play in the evaluating process?
- Do implementing organizations or donors ask for a follow-up of your work (e.g. based off of recommendations given in the report)?

Implementing Organizations

Ideally, evaluations should serve as a tool to help organizations identify best practices and enhance their programs. However, it is not clear if this lofty goal of evaluation is realized in practice. Therefore, it is important to ask organizations to clarify the role that evaluation plays in their projects and what they expect to get out of the evaluation process. In this way, we can see not only what is learned through evaluation, but also ask if the knowledge gained was useful to the organizations in question.

- What do you learn from external evaluation?
- Who learns from evaluation? Does learning vary at the different levels of the organization (e.g. project team, board, members, partners, etc.)?
- Does the knowledge acquired through external evaluation of your projects lead to changes in the way you develop or conduct your projects? If so, how?
- What are the barriers to changing how projects are conducted?
- How do you participate in external evaluation?
- Are beneficiaries of your program or project involved in the evaluation process?

ISSUE 3: Innovation and Evaluation

Donors of development projects and programs

Classic ex-post external evaluations are getting challenged with new criticisms that lead to the elaboration of new practices. In order to assess how evaluation practices might be changing, it is important to understand how donors view innovative evaluation practices, and how these practices influence funding decisions.

- Do you see evidence of innovation in the evaluation field? Are you using new techniques?
- What are the potential risks or benefits of funding projects that use non-traditional evaluation techniques?
- Can you identify shortcomings you see in current evaluation practices?
- Are there ways to overcome these shortcomings to make funding projects more effective?
- Do reciprocal evaluations, in which program beneficiaries evaluate donors, tend to foster learning?

Evaluators

In order to assess how evaluation practices might be changing, it is important to understand how evaluators view innovative evaluation practices. Because of the direct role evaluators play in assessing project outcomes, they are also uniquely suited to identify both strengths and weaknesses in current evaluation practices. The challenges that evaluators face in their daily work could provide an entry point to identifying opportunities for innovation.

- Can you identify shortcomings in current evaluation practices?
- Are there ways to overcome these shortcomings to make evaluating projects more effective?
- Do you see evidence of innovation in the evaluation field? Do you use new techniques in your work?
- Is there an exchange of good practices within regional or international evaluation networks? Between different stakeholders within platforms dedicated to evaluation such as F3E?

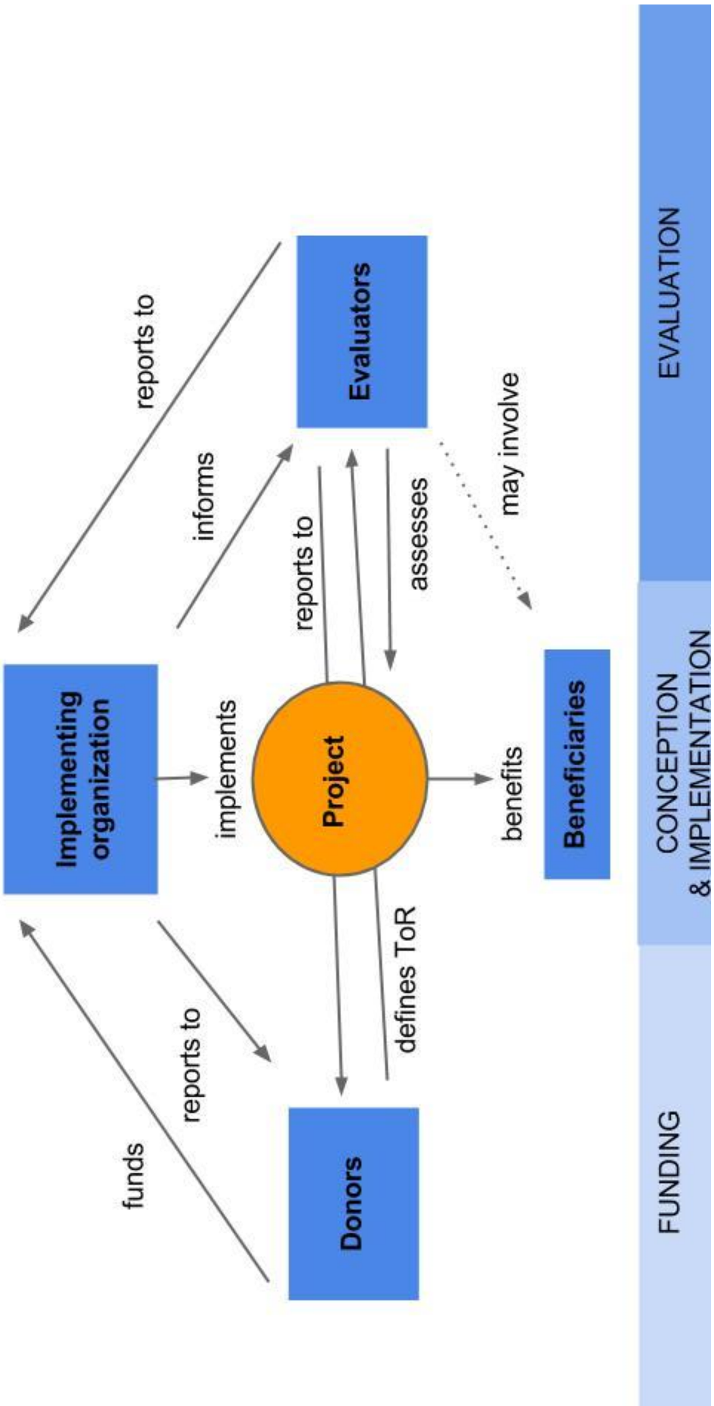
Implementing Organizations

In order to assess how evaluation practices might be changing, it is important to understand how organizations view innovative evaluation practices, and how these practices influence their programs.

- Can you identify shortcomings in current evaluation practices?
- Are there ways to overcome these shortcomings to make projects more effective?
- Are you willing and see it possible to experiment new approaches while conducting the projects? If yes, for what objectives? What ways do you see to shape the investment of the donors? Do you think evaluation results could help you with that?
- Do you use any innovative evaluation techniques while conducting internal evaluation? (if you conduct it)

Annex 2: Diagram of Stakeholders' Interactions

To bridge our deskwork and our interview process we developed a diagram to illustrate the interactions between different stakeholders in the evaluation process. The diagram shows the possible flows of communication between the various actors throughout the different parts of a project cycle.



Annex 3: Interviewees

When we began conducting interviews, our work was focused on developing a regional understanding of current evaluation trends and norms. Thus, we organized our interviews by region. We sought to have at least two representatives of each of our five original regions (Francophone Europe/ North America; Anglophone Europe/ North America/ Pacific; Latin America; Anglophone Africa/Asia; Francophone Africa/Asia). Interviews were conducting in person when possible, however, due to the international nature of our work, many interviews were conducted at a distance. We communicated via telephone and Skype.

Interviewee	Region (language)	Organization
Sue Soal	Asia/Africa (EN)	CDRA
James Taylor	Asia/Africa (EN)	CDRA
Ian Hopwood	Asia/Africa (EN)	Save the Children/UNICEF
John Healy	Asia/Africa (EN)	Independent consultant & former director of the Atlantic foundations
Oumoul Khayri Ba Tall	Asia/Africa (FR)	OKT-Consult
Emilie Aberlen	Europe (FR)	AFD
Marc Totté	Europe (FR)	Inter-Monde (ONG)
Patrice Dufour	Europe (FR)	World Bank/Aid Impact
Florent Bedecarrats	Europe (FR)	AFD
Evaluator from a private French evaluation firm	Europe (FR & EN)	Private evaluation firm
Jan Van Ongevalle	Europe/USA/Pacific (EN)	Ku Leuven
Mary Hall	Europe/USA/Pacific (EN)	EU monitor
Luis Soberón	Latin America (SP)	ELAC Project
Sergio Martinic	Latin America (SP)	
Pablo Rodriguez Bilella	Latin America (SP & PO)	EvalPartners

Annex 4: Interview guide

Interviewee
Organization
Date interviewed by interviewee

ISSUE 1: Defining Evaluation

- For you, what is evaluation?
- What is your main goal as an evaluator?
- How is this goal determined? (By donors, organizations that implement the project, personally identified needs?)
- In your opinion, if and to what extent could evaluation be used as a tool for social change?
- In your work, have you noticed any regional differences in approaches to or methods of evaluation?

ISSUE 2: Evaluation and learning

- What do you seek to learn from evaluation?
- Who learns from evaluation? Who in your opinion should learn from evaluation?
- What do you do with the results of evaluation? How do you publish it?
- Does the knowledge acquired through evaluation lead to changes in how evaluations are conducted (demands from donors and your own practices)? If so, how? If not, what do you think are the barriers to changing evaluation practices?
- To what extent do you involve other stakeholders in the evaluation procedure (donors, communities)? Are you able to communicate with other stakeholders during the evaluation process? How does this affect learning (your learning, donors' learning and the learning of the beneficiaries)?
- Is there a follow-up with your other donors or the evaluation team? If so, what form does it take? Do you think that donors and organizations are learning through your evaluations?

ISSUE 3: Innovation and Evaluation

- Can you identify shortcomings in current evaluation practices?
- Are there ways to overcome these shortcomings to make evaluating projects more effective?
- Do you use new techniques?
- Is there an exchange of best practice within the evaluation network?
- General Comments