Measuring International NGO Agency-Level Results

Carlisle J. Levine, Ph.D., BLE Solutions, LLC

Tosca Bruno van Vijfeijken

Sherine Jayawickrama

May 2016

InterAction

ACKNOWLEDGEMENTS

BLE Solutions

TABLE OF CONTENTS

LIST OF CHARTS

I. Background

International non-governmental organizations (NGOs) work in numerous countries and are engaged in many sectors, providing services, building capacities, strengthening systems, and influencing policy and practice. While they are increasingly able to provide evidence of the effectiveness of their project and program-level investments, they are still figuring out how to demonstrate the difference they are making as organizations. In response, many international NGOs create agency-level measurement systems, with varying degrees of success.

These agency-level measurement systems take a variety of forms, described in more detail below. Some aggregate indicator measurements in priority programming areas. Others report on indicators aligned with an agency's strategic plan, theory of change, program priorities, or mission. To complement indicator measurement approaches or as an alternate approach, some gather findings from meta-evaluations or samples of representative evaluations, or they conduct new evaluations to assess impact.[1]

To examine whether or not the building of agency-level measurement systems is a

---

1 In this paper, "system" refers to the overall approach towards agency-level measurement that an NGO chooses, not just the IT systems that support the approach.

Chart 1: Approaches featured in organizations' agency-level measurement systems (N=71)
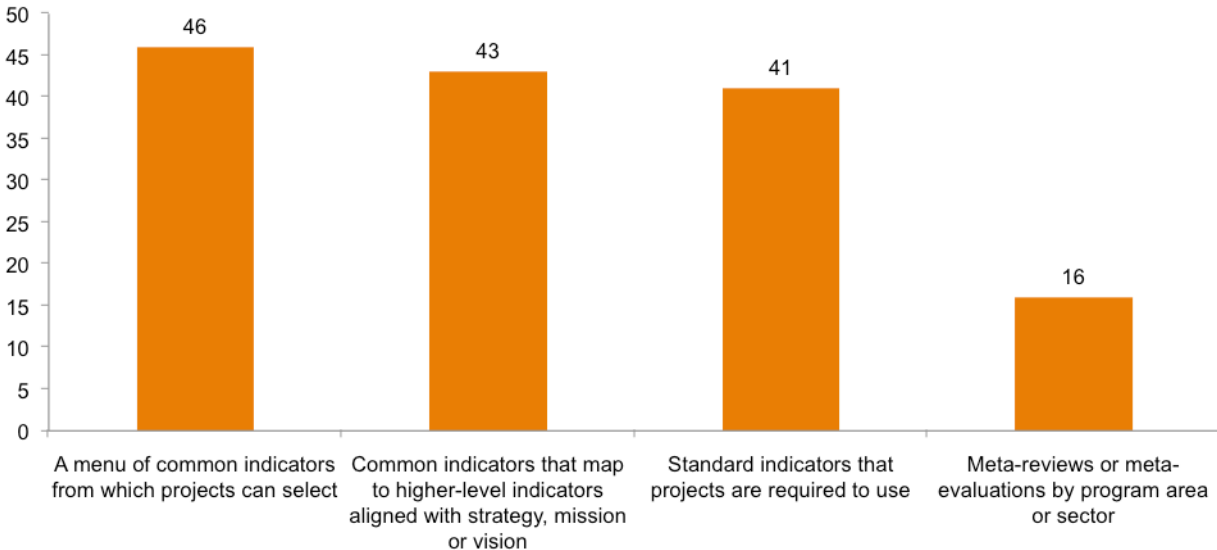
Note: Survey respondents listed all that applied.

worthwhile endeavor, under what conditions it delivers benefits and what are its potential challenges, NGOs wanted to learn from each other's experiences. Eleven international NGOs as part of InterAction's Evaluation and Program Effectiveness Working Group (EPEWG) commissioned this study to enhance their understanding of these systems, what works, and for what purposes.

This white paper and its accompanying brief draw on the existing literature on the topic, broad input from international NGOs (through an EPEWG focus group, a survey[2] primarily with headquarters-based staff working for U.S.- and European-based NGOs, and a world café with InterAction member CEOs), and deep

exploration of the experiences of 17 InterAction member organizations (through case studies and interviews).[3]

The paper describes motivations for creating such systems, the expectations and assumptions associated with them, and the nature of the systems. It includes three brief cases as examples. The paper then analyzes what it takes to build and maintain them, their use, key challenges, benefits, risks, trade-offs, and costs. Based on that analysis, the paper offers a series of recommendations to help NGOs decide whether or not agency-level measurement makes sense for them, and, if so, how to develop systems that meet their needs.

---

2 To encourage candor, we allowed anonymous survey responses. However, therefore, we do not know the number of organizations represented by the individual responses, nor the number of individuals per organization who responded. As a result, we do not know if the data reported are skewed toward the experiences of a small number of organizations. We have kept that in mind in our analysis. For example, when there are discrepancies between survey findings and case study or interview findings, we believe the case study and interview findings have more credibility.

3 When referring to survey responses, the paper provides the exact number or percentages of respondents. When referring to the case studies, world café, focus group or interviews, which do not lend themselves to precise numbers, the paper categorizes responses as relevant to: the vast majority (close to 100%), most (more than 60%), many (40-60%), some (20 -40%) or a few (two or three) respondents.

II. Overview of Agency-Level Measurement Systems

A. Why do NGOs adopt agency-level measurement systems?

International NGOs have a variety of motivations for adopting agency-level measurement systems. Those motivations are captured here. How they play out in reality is described in later sections.

NGOs experience **internal pressures** from their boards, leadership, and staff to demonstrate agency-level results. To get an overall picture of organizational results, boards and senior leadership teams often request (or are provided with) dashboards containing high-level data about an organization's reach, activities, and accomplishments. For some board members, being able to produce such a dashboard is evidence of an international NGO's sophistication.

Better understanding organizational reach and results, and being better able to market achievements are the primary reasons international NGOs develop agency-level measurement systems. NGOs also cite a desire to enhance their accountability to donors, whether institutional donors or individuals.

Leaders and staff want to know what they are accomplishing on an organizational level for a variety of internal organizational reasons.

> Just the act of adopting an agency-level measurement system makes us better at what we do. It makes us focus more on results and learning, and it encourages us to invest in M&E capacity.
>
> *– CEO of an international NGO*

At minimum, they want to know how many people they are reaching with what kinds of activities.

Are we contributing to higher goals? When focusing on higher goals, most want to gauge progress against organizational strategy and test their organizational theory of change, according to the 17 case studies. Some place this assessment within the context of the overarching agenda of the development community: using the data to identify their strengths and to help assess their organization's contributions to broader community goals. A few want to measure their progress toward achieving their organization's mission. Their findings may confirm their established pathways or point to new strategic directions.

In a few NGOs involved in this study, the desire to understand organizational effectiveness has been so strong at the country or regional level that absent an agency-level system, some regional and

Chart 2: Motivations for developing an agency-level measurement system (N=66)
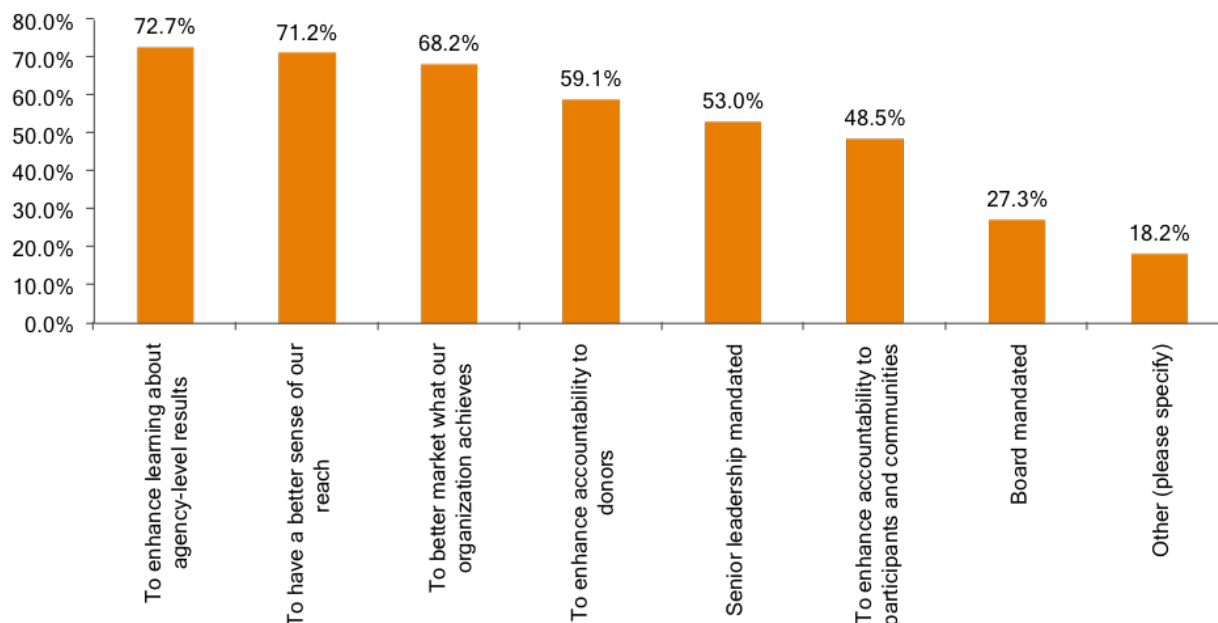
Note: Survey respondents listed all that applied.

country offices have developed their own results measurement systems.

How do projects contribute to an organization's goals and vision? Many organizations want to better understand how projects fit together to contribute to goals beyond the project level; this was seen in the case studies submitted. For some, this type of review helps break down organizational silos. For at least two NGOs, it supports efforts to unify an international federation or confederation.

In terms of **external pressures**, some NGOs believe donors, rating agencies, and the public want information on results at an organizational level. In the past, some donors invested in international NGOs' overall capacities (e.g., USAID Private Voluntary Cooperation grants and the Programme

Partnership Arrangements of the U.K. Department for International Development), and therefore may have been interested in agency-level results. Today, donors confirm that they are interested in evidence of reach and results, and rating agencies join them in valuing a commitment to demonstrating results. However, donors most often request results at a sector or country level, where they currently tend to focus, and rating agencies mainly seek evidence of results measurement at the project (not organizational) level.

Transparency and accountability. Some NGOs believe that sharing the data from an agency-level measurement system is important to demonstrating transparency to partners, program participants, donors, and the general public. These systems can provide basic information about presence, reach, and activities, as well as results.

Influence. Some NGOs also believe agency-level results data can strengthen their ability to influence government policies and practices. Data on presence and reach provide a sense of an organization's constituency, while results data can demonstrate what works under what circumstances.

International NGOs tend to decide to create agency-level measurement systems based on the information they hope the systems will provide, rather than first considering the resources they have available to invest in those systems. Only a few of the 17 NGOs in this study described explicitly how they considered the resources required to develop and maintain such a system prior to making an investment decision. All were clear about the staff hours and budget they would allocate to the task, as well as the timeframe in which they wanted the system to produce data. However, these factors influenced the system's scope, not the decision about whether to build a system at all. In almost every case, the actual resources and timeframe were underestimated, as discussed below.

B. Expectations and assumptions

There are often differing expectations about the type of information the system should provide. Higher-level decision makers tend to want a system with key information, often based on aggregated data they can use to take an organization's temperature and make executive-level decisions on strategy adaption and resource reallocation, according to the 17 case studies submitted. In contrast, technical and program staff might want greater levels of

specificity that they find more meaningful and that can help inform program decision making. Marketing and communications staff often want data that help position the organization favorably in a competitive landscape.

> No matter how much the system was intended for internal use, the pressure to also use the data for external communication is unavoidable.
>
> *– NGO staff member*

Organizations often underestimate what it will take to achieve their desired system, as evident in the case studies submitted. Organizations underestimate the time, funding, expertise, and on-going technical assistance required to develop and maintain a useful management information system. They also tend to overestimate the project management, data collection and analysis, and broader monitoring and evaluation capacities of their staff; this can affect the organization's ability to capture quality data and use it appropriately. Finally, organizations tend to underestimate the change management capacities required and the organizational culture change that may be needed to implement such a system. The success of agency-level measurement systems depends on having a culture that values data and evidence, as well as staff that see themselves working in the interest of the broader organization (and not just in the interest of their project, sector, or country).

How information from the system will be used can be a point of contention. Some systems are explicitly designed for use by the organization's leadership and for external use. Others are designed for marketing or

fundraising. Others claim to be designed primarily for project and country program staff. Still others are expected to be used by all staff. However, even if designed for project and country program staff, organizations have found that the pressure to use data for marketing, communication, and fundraising can be overwhelming. And in some cases, pressure to produce information for marketing and communications has affected the ability of the system to produce information of value to senior decision makers and program staff.

Very few systems manage to simultaneously produce data specific enough for program managers and technical staff and also general enough for forms of agency-wide aggregation that is useful to executive leadership and marketing and fundraising staff. Broad system buy-in (which is difficult to achieve) and data use depend on the relevance, quality, and timeliness of the data for a diverse group of organizational stakeholders.

C. What do agency-level measurement systems look like?

This section covers both the nature of agency-level measurement systems and the role data management platforms play in supporting them. It concludes with three brief case studies covering the development of actual agency-level measurement systems.

Designing agency-level measurement systems

The vast majority of international NGOs that submitted case studies designed their **agency-level measurement systems in one of two ways: top-down or bottom-up,** with some using both approaches for two different aspects of their systems:

1. **Top-down:** Some NGOs started by identifying organizational-level measures, often aligned with their strategic plans or mission-level themes. They then asked different parts of their organizations to track indicators associated with these measures, or they identified indicators already being used by different parts of the organization that were relevant to these measures. Alternatively, they selected topics of interest at the agency level and conducted meta-evaluations, thematic evaluations, or impact evaluations focused on them.

2. **Bottom-up:** Other NGOs started with systems that could aggregate data gathered in country programs and elsewhere in order to make organization-wide statements about results. Some of these systems started with project-level indicators that were either already standardized or that could be standardized and that were commonly measured across country programs. Others started by reviewing existing project, program, or sector evaluations and selecting a representative sample of them as evidence of what the NGO was achieving.

In both top-down and bottom-up approaches, the indicators provide a sense of breadth, while the evaluations provide context and depth. In one case, the NGO focused primarily on meta-evaluations and thematic evaluations to gain information about agency-level

performance, although they also gathered quantitative information about reach. These findings were echoed in the survey results, as shown earlier, and in a similar study[4] that added a variation on the top-down approach: using basket indicators,[5] which allowed sectors and projects to keep their more specific and, therefore, meaningful measures.

There are benefits and challenges to both top-down and bottom-up approaches. Top-down systems prioritize measures and information that reflect organizational-level priorities, such as those in an organization's strategy. However, in such systems capturing relevant data without placing too much additional burden on staff is challenging. Moreover, unless there is strong alignment between agency-level strategy and country-level strategies, country program staff cannot always relate to the agency-level measures or information demands.

---

4 Nigel Simester, "Summarising Portfolio Change: Results Frameworks at the Organisational Level," M&E Paper 10, INTRAC, January 2016, http://intrac.org/data/files/resources/888/Summarising-portfolio-change-results-frameworks-at-organisational-level-FINAL.pdf.

5 "Locally defined indicators, while not themselves comparable across projects, might fit within common categories of indicators [or baskets] that might lend themselves to comparison. For example, a locally defined indicator of increased trust might be 'mutual use of a road between the two participating communities increases.' This indicator would not be comparable across projects. However, if it is included within a common category of peacebuilding indicators [or a basket indicator], such as 'increased mobility,' then the results it describes might be both comparable and able to be meaningfully aggregated to describe results at a [country], regional or agency level." Carlisle Levine, Catholic Relief Services' (CRS) Guidance for Developing Logical and Results Frameworks, Catholic Relief Services, 2007, http://betterevaluation.org/resource/guide/crs_develop_logical_results_framework.

Bottom-up systems minimize the added burden on staff, but face challenges in aggregating data in a way that takes context into consideration and allows the measures to remain meaningful. Bottom-up systems may also generate hundreds of indicators, which can entail ongoing measurement needs the organization cannot sustain. Using a basket indicator approach that creates broad categories of indicators can mitigate these two challenges, although it increases the burden on those analyzing the data. Finally, determining criteria for selecting a representative sample of project, program, or sector-level evaluations can be challenging and subject to critique after the fact.

NGOs frequently use agency-level measurement systems to help measure progress against their organizational strategies, as seen in the top-down approach. While some systems are results-focused, many also include measures focused on operations and program quality. In a few of the case studies, the organizations tried to test their theories of change, and a couple framed their systems within their organizational missions. Bottom-up systems sometimes focus on program priorities, while others try to add up all that an organization is accomplishing in all its areas of work. Only one system in the case studies focused exclusively on counting numbers of people reached.

Many agency-level measurement systems end up focusing on outputs, at least in their early days (whether or not they were intended to do so). This is because outputs are easier to standardize and demand less additional work by country offices. At the same time, there is concern about

> What does "number of people trained" tell me? If one person participates in a one-day handwashing training in one country, and another person participates in a year-long, multifaceted, conflict mitigation training in another, do they add up to two?
>
> *– NGO staff member*

how meaningful these outputs are once aggregated. For example, with the indicator "number of people trained," does the number reported include people who received the same length and depth of training, focused on similar numbers of topics? Aggregated outputs are only meaningful if clearly defined. They are more meaningful if their links to outcomes have already been established.

Many agency-level measurement systems find it necessary to include a hybrid of standardized indicators (sometimes outputs, sometimes outcomes) and qualitative data that provide context or evaluations that better get at the achievement of outcomes, as described above. In this design, international NGOs recognize that indicators provide only a partial story and must be augmented with additional information to create a well-rounded picture and deep understanding of what an organization is achieving. Organizations use meta-evaluations, systematic reviews, impact reports, thematic evaluations, project evaluations, and impact evaluations to complement their standardized indicator measures.

International NGOs that have not developed agency-level measurement

systems or that have developed limited systems have found other ways to assess their work beyond the project level. They use the approaches described above. They also collect country-level data that they compare within a country over time, or sector-level data that they compare across countries and time. However, even in organizations that are far from having the resources and capacities to implement an agency-level measurement system, senior leadership and boards, as well as business development teams, expect that in future years, the NGOs will be able to present program quality and/ or effectiveness information in a summarized, organizational-level manner.

Data management platforms

International NGOs agree that data management platforms need to support agency-level measurement systems, but cannot and should not determine their parameters. For NGOs, understanding the human side of the equation is paramount. This turns on understanding staff realities and constraints – especially at the country level, and designing a data management platform that responds to that. This requires close collaboration between monitoring and evaluation (M&E) and information technology (IT) staff, as well as good communication with those who will be charged with data collection, particularly within country programs. According to one case study, having an IT manager with deep country experience is an asset.

Nonetheless, while they should not drive an agency-level measurement system, **data management platforms are critical for**

their success, as seen clearly in the case studies. They determine who has access to the data and how easily they can access and use the data – important parts of the incentive structure. These two factors have tremendous influence over staff buy-in into a measurement system and their willingness to take the time both to ensure that it contains quality, up-to-date data and to use it to inform decision making.

Budget considerations often drive the selection of data management platforms. In some of the contributed cases, the data

management platform had no allocated budget; in others, NGOs were investing millions of dollars in their development. Some organizations were able to create satisfactory solutions that only cost $20-30,000.

An allocated budget to develop the data management platform offered more possibilities in determining its best option, but there was no consistent, direct relationship between dollars spent and success of the data management platform. In at least three cases, poorly performing purchased software had significant negative

The data management platforms selected included Excel, off-the-shelf database software, and custom-made database software. According to the international NGOs, each has pros and cons.

- **Excel** is already on staff computers, thus avoiding an additional cost and the problem of poor internet connectivity (except when emailing content from one computer to another). Staff are familiar with Excel, minimizing the need for training and IT assistance. It offers data analysis and presentation options. However, using it becomes difficult when managing large amounts of data from many sources. It can also be hard to standardize spreadsheets for diverse projects, requiring a large team to modify and update them.

- **Off-the-shelf options** come in two forms: those primarily designed to manage project results data (e.g., District Health Information Software (DHIS)); and those that are customized for the purpose (e.g., Raiser's Edge). Systems, such as DHIS, may be well-established and in broad use by entities engaged in development, which means that the designers may be more open to modifying it to meet users' needs. Systems such as Raiser's Edge are often offered to international NGOs at reasonable prices and/or might already be used by their organizations for non-program functions. However, because these systems were never designed to manage project results data, it is not easy to modify them to meet agency-level measurement systems' needs.

- **Custom-made systems** offer the promise of being most responsive to an agency-level measurement system's needs. However, the success of these systems seems to depend heavily on the availability of ongoing IT support to make adjustments as the needs of the organization change and offer technical assistance. They also do not benefit from the learning that informs updates to off-the-shelf systems, although such updates can be built into contractual arrangements.

effects on the systems' use. In two of those cases, shifting from purchased database software to Excel improved the system's usability and actual use. However, neither NGO believed that Excel would be a viable long-term option. Those who expressed the greatest satisfaction with their system had budgets in the ten thousands, rather than in the millions of dollars.

All agreed that to foster buy-in into the agency-level measurement system, the data management platform needed to be user-friendly. This meant that it had to function on- and off-line. All staff members needed access to the data. The data needed to be easy to upload and easy to use. The system needed to communicate with other systems at the global, country, and regional levels, so that it would be a time saver and not create a requirement for double reporting. Most importantly, the NGO needed easy access to ongoing information technology support to facilitate the system's use. In addition, some staff needed the assistance of M&E specialists to make sense of the data.

Case studies

The following three cases present examples of how an agency-level measurement system can be designed to promote use. The first case, Helen Keller International, shows a system focused more on process measures. The second case, CARE International, includes basic project data and outcome information. The third case, Bread for the World, describes one of the longest existing systems – a system that includes data covering activities to outcomes, as well as operations.

Helen Keller International

Helen Keller International (HKI) works to save the sight and improve the nutrition, health, and lives of the most vulnerable and disadvantaged in 21 African and Asian countries, as well as in the United States, through more than 120 programs. HKI's programs are geographically diverse, cover a number of sectors, use different approaches, and vary in scope and size.

Motivations: Helen Keller International developed its agency-level measurement system to measure its progress on a global and country level against its global strategic plan for 2010-2015 and the country strategic plans that are aligned with it, although adapted to local contexts.

Expectations and Assumptions: The need for common metrics became apparent during country-level strategic planning workshops; to draw any global conclusions based on country-level activities, common metrics generating compatible data were required. The system had to be meaningful, but also sufficiently light, given other demands on staff's time.

The Nature of the System: The system, which has been gathering data since the 2012 fiscal year, consists of 28 process measure indicators, defined to measure compliance with program standards for program quality, gender equity, behavior change, and M&E systems; partner feedback; and performance on 14 strategies associated with the three program objectives of HKI's strategic plan. The strategic objectives relate to ensuring in-country and regional capacity, developing and implementing replicable and evidence-based

models, and creating a supportive climate or enabling environment for service delivery. Defining these indicators was an iterative and inclusive process. They complement a longer-standing list of shared indicators that mainly focus on program outputs and HKI's organizational effectiveness standards. Now with the development of HKI's strategic metrics, HKI is bringing its measurement systems together and making the data more consistent and accessible. In the future, HKI will add metrics related to organizational operations.

Use and Satisfaction: The data help HKI's executive team and board of trustees monitor agency progress on its strategy. The trustees are satisfied with the user-friendly presentation of the information and its utility in helping them track HKI's progress. The data are included in HKI's reporting on organizational effectiveness for both internal and external purposes. Managers are generally positive about the system. They use the information at strategic decision-making points, such as during annual work planning and mid-year reviews. The program standards serve as guidance for new program design. However, qualitative information is also needed to provide context. HKI collects specific examples to illustrate its accomplishments related to each indicator or performance standard.

Costs and Benefits: The system has provided concrete data on organizational performance, especially as related to HKI's strategy. Reflecting on the data produced has informed the development of new programs. Accountability to the board has improved, as have discussions between the board and senior management about HKI's work. In part

thanks to the system, HKI won the 2014 Kravis Prize for Nonprofit Leadership for its efforts to measure organizational effectiveness. The costs of the system, which have been primarily staff time and opportunity costs, have been worth the benefits achieved, according to HKI.

Lessons Learned

1. Use an iterative and inclusive process to ensure buy-in.

2. Design a system that takes into account the resources (staff, skills, financial) available to it.

CARE International

CARE International is a global confederation of 14 member organizations. In 2015, CARE worked in 95 countries around the world, supporting 890 poverty-fighting development and humanitarian aid projects to reach more than 65 million people. CARE's actions work to deliver lasting change in the lives of poor and vulnerable people, particularly women and girls.

Motivations: In 2012, CARE International began to develop the Project/Program Information and Impact Reporting System (PIIRS) in response to an organizational need to develop a joint platform for all CARE offices, dedicated to collecting, accessing, and reporting relevant information on the work CARE does globally and the breadth and depth of changes to which this work contributes. The goal was to improve CARE's organizational ability to communicate

and learn from its work and evidence its contribution to significant changes in a unified, more coherent and more accurate fashion. CARE had a number of systems in the past, but this was the first one to cover all locations where CARE works and also respond to strategic priorities agreed by the confederation.

Expectations and Assumptions: PIIRS' content and tools had to fully align with CARE International's programmatic priorities. It had to build on existing experiences and be cost-effective in its design and implementation. It was believed that having PIIRS in place would improve accessibility to basic, standardized information on the breadth and depth of CARE's work globally, which could then be shared internally, with CARE's board and donors, and with the general public.

The Nature of the System: PIIRS is structured around two main tasks. The first is generating general information about CARE's projects and programs around the world in order to show the breadth of CARE's work (four years of data collection undertaken to date). The second task, for which the processes are still in the design stage, is generating impact information from CARE's projects and programs around the world, based on standardized global outcome and impact indicators related to CARE's programmatic priorities, in order to demonstrate the depth of CARE's work.

All the information is hosted in an online information system developed using the previous experience of a country office. This customized system is simple and modest in price. Significantly, the developer is available to provide ongoing technical assistance and make adjustments to the system as CARE's priorities and needs evolve.

Use and Satisfaction: So far, CARE staff use the information for learning, analyzing sectoral performance, developing proposals, and communications. The intensity of the data's use greatly varies from country to country or between different teams. The data are also used as input to impact reports, speaking to CARE's contribution to significant changes in specific areas of work.

Having a unified and accessible platform with data about the scope of CARE's work worldwide is recognized as a big accomplishment, although it requires constant dialogue around the level of detail of the information collected. As in any information system, there is definitely room for improvement concerning data quality. Increasing the quality of both general and impact information requires capacity building and improvements in M&E practices so projects and programs can better capture evidence of CARE's contribution to change.

Costs and Benefits: Having a single project information platform has been an important breakthrough in the CARE confederation. Establishing global indicators fully aligned with CARE's Global Program Strategy is also an important step toward demonstrating impact and improves CARE's organizational coherence in its most strategic areas of work. Costs are related to the online database and staff time, and investments still need to be made in staff capacity building and revised M&E guidance. For now, benefits outweigh costs.

Bread for the World

Bread for the World advocates for U.S. policy changes to help end hunger in the United States and globally. It produces research and policy papers, uses traditional and social media, mobilizes individuals and churches as advocates, and engages in direct advocacy with the U.S. Congress and administration.

*Motivations***:** As Bread for the World received increasing numbers of foundation grants, those foundations asked Bread questions about its results and impact. Individual donors were also interested. Bread wanted to develop a tool to show that it takes results seriously. Internally, Bread wanted to know its strengths: both its reach and the policy change it was influencing.

*Expectations and Assumptions***:** Bread for the World's leadership expected that its agency-level measurement system could

provide a snapshot of data showing its reach and what difference it was making. This analysis would be updated annually, with a dashboard with key monitoring indicators tracked every six months.

*The Nature of the System***:** Bread for the World's agency-level measurement system was designed in 2007 and updated in 2014 to align with a new agency strategy. The indicators measure its progress against its strategy and theory of change. In this way, different departments can see their contributions to Bread's overall achievements. Bread also collects stories describing its contribution to policy change objectives. This combination of indicators and stories provides both breadth and depth.

Bread for the World uses three information management systems to track its data: financial management software, communications software, and Raiser's Edge to capture all other activities. These systems do not communicate with each other. Raiser's Edge, a relationship management and fundraising software, is an imperfect solution for this broader use. However, Bread has not yet identified more satisfactory software.

*Use and Satisfaction***:** For accountability, Bread for the World presents its dashboard of key monitoring indicators to its board annually. The board is most satisfied with the data. They discuss them and ask good questions. Bread's management team uses the key monitoring indicators and a larger set of data to raise questions about performance, identify investments required to better achieve its targets, and set goals for upcoming years. The management team is satisfied once the data

are available, since the data inform strategic decision making. However, management and staff continue debating definitions associated with measures, and Raiser's Edge remains a cause of frustration. Individual departments are less likely to use the system. For most, the data are not collected frequently enough to inform decision making. Also, for ease of use, the evaluation metrics do not capture all departmental goals. For staff who do not use the data in their day-to-day work, collecting data can be a burden. Bread increasingly uses metrics, testing, and benchmarks to inform its digital advocacy. Management anticipates using the evaluation more frequently to inform "real time" decision making. This is driving more consensus on definitions, more uniformity of data entry practices, and increased ease of accessing reports.

Costs and Benefits: Overall, the system is worth it. Bread staff learn things about themselves that they would not otherwise know. The system provides hard numbers to back up conversations about required improvement, taking the emotion out of it. As one staff member noted, "When you look at numbers collectively, reflecting on all of us, it facilitates the conversation about what we do well and what we don't do so well."

Lessons Learned

1. Be clear about what you need to know at the organizational level.

2. Have a theory of change with associated metrics linked to vision, mission, and strategy.

3. To be perceived as useful, have the system complement other M&E processes and relate to mission critical work.

4. Include data collectors in identifying the right measures in order to get their buy-in.

III. Findings

A. Laying the groundwork for system success

Findings from this study make clear that success depends on more than just technical design. Many participants noted the importance of certain organizational factors, some of which should be in place before design even begins. This section covers those factors. All of them depend on first determining if an organization is in the midst of any fundamental overhauls of its strategy and vision; that topic is covered below in the Lessons Learned section.

International NGOs recognize the importance of having a culture that values data-driven decision making to the success of agency-level measurement systems. In such organizations, staff value data and are more likely to provide quality and timely data and use it to inform decision making.[6] However, in many international NGOs, creating such a culture is still very much a work in progress.

Having a consistent champion within an organization's executive leadership bolsters

Organizational readiness for an agency-level measurement system depends on having:

- An organizational learning culture that values data-driven decision making;

- A champion within the executive leadership who invests time in it;

- A clear organizational strategy, theory of change, and/or priorities;

- A clear audience and purpose; and

- Adequate skills for data collection, management, analysis, and use.

a system's success. In at least three case studies, NGOs said having such a champion greatly enhanced organizational buy-in and system use. Critically, this champion must be outside of the monitoring, evaluation, and learning function. She or he needs to have time to devote to this effort, understand the options, needs, constraints, costs and benefits, and also take time to analyze and really understand and use the data.

In two cases, NGOs abandoned systems in part because the original champion left the organization, leaving the system without an advocate at the top. In other cases, while system development and implementation are making progress, progress may be

---

6 For more on this topic, see Laia Griñó, Carlisle Levine, Stephen Porter and Gareth Roberts, *Embracing Evaluative Thinking for Better Outcomes: Four NGO Case Studies*, InterAction and CLEAR-Anglophone Africa, June 2014, http://www.interaction.org/node/3818.

incremental, with those managing facing challenges in getting broad buy-in.

A clear organizational strategy, theory of change, and/or priorities tell an organization what is important to measure at an agency level. In some cases, the linkage between these and an agency-level measurement system created buy-in into the system. Staff find an agency-level measurement system more useful when it is integrated into larger and ongoing processes or frameworks. With that linkage, in at least one case, the system also helped reinforce an organization's strategy goals. In other cases, with less clear linkage between strategy and the system, staff were less convinced of the system's relevance and value.

Clarifying who will use the system and its purpose is also important. In cases lacking clarity regarding the system's users and purpose or that tried to be too many things to too many people, satisfaction and use suffered. When the users and purpose were clear from the beginning, expectations about whom the system would serve and how were more realistic, and satisfaction and use were both higher.

Staff need adequate skills in basic project management, data collection, management, and monitoring. In many cases, NGOs found that these skills were lacking or uneven either prior to or in the process of developing a system. Some NGOs used a gap analysis to identify existing evaluation capacities and integrate additional capacity building into the core of the agency-level measurement systems. In other cases, organizations have hoped

that as staff become more practiced at collecting, managing, and using data, their skills will improve and, along with them, the quality of the data they provide and their use of the system. In one case, an organization decided to not develop a traditional agency-level measurement system until it was able to improve its staff's capacities.

B. Agency-level measurement data: who uses it and how?

Although many of the systems explored in this study are still being developed or are relatively new, most are being used. Of the 17 organizations that submitted case studies, 12 reported system use. Five reported that it is too early to report on use, while three of the 12 reported that parts of their systems were not (yet) in use.

Chart 3: Case study NGOs reporting system use (N=17)
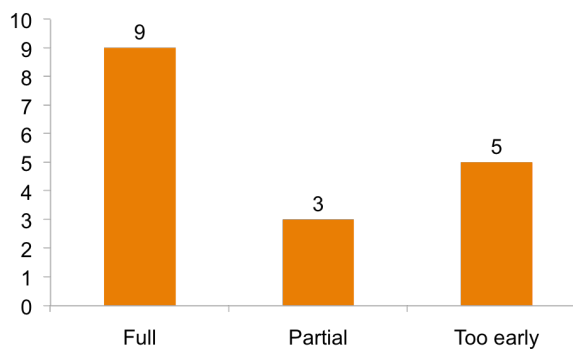
Intended users and satisfaction

Most agency-level measurement systems are requested by an organization's leadership, are designed to meet their needs, and are intended for their use. Senior leadership was by far the most

frequently listed intended group of users by survey respondents (89%). Donors, boards of directors, marketing and communications staff, and program staff were also listed by a majority of respondents (65-68%). Most of these groups of intended users indeed turn out to be the systems' primary users. However, although case study contributors often listed regional and country-level staff as among the primary intended users, they found that in actuality, these staff members used the data less than anticipated.

Board members and senior leadership are most likely to express satisfaction with agency-level measurement systems, explaining that the systems provide useful information that helps them track progress and raise questions, although a few raised questions about how measures were aggregated. Evaluation staff are also likely to be satisfied with their systems, as

> The more we [generalized] the measurement we were seeking at the agency level, the more it became clear such measurements were [contextual] because what you can say about collective contribution to the mission is actually very localized.
>
> *– NGO board member*

demonstrated in the survey results. Country-level and other frontline staff report the lowest level of satisfaction, saying that the data do not accurately reflect context and accomplishments. In one case, a system seemed to fall between the cracks, not satisfying senior leadership and board or frontline staff. This happened because the agency did not make a clear choice about primary users and ownership, and resulted in the non-continuation of the system.

Chart 4: Main intended user groups for agency-level measurement system data (N=64)
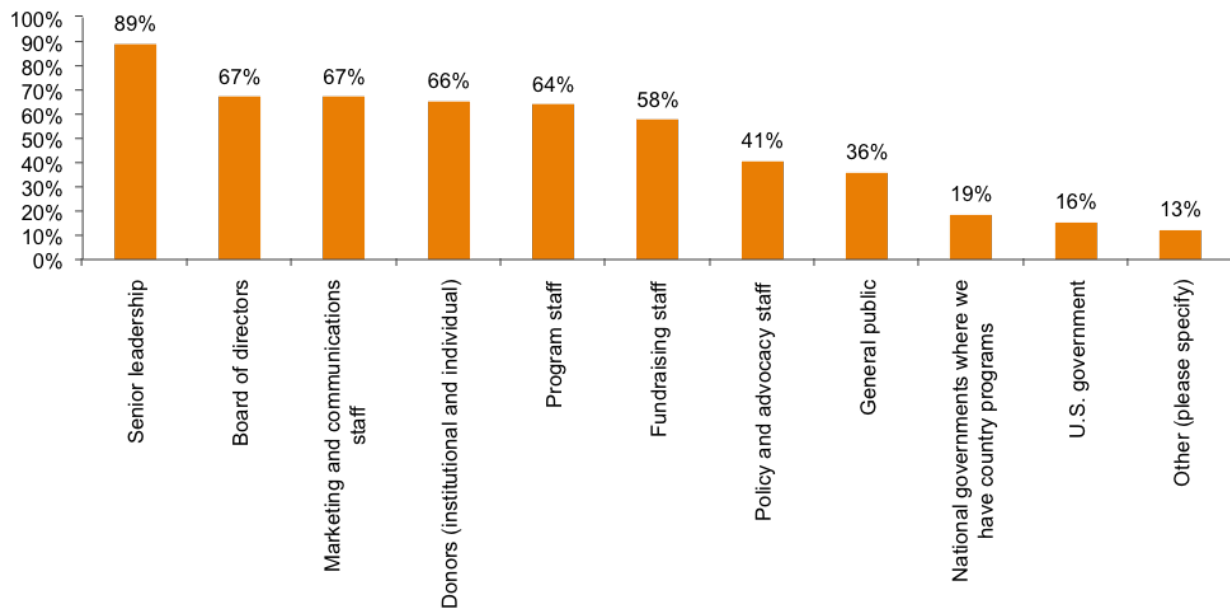
The satisfaction of field staff with agency-level measurement systems seems to reflect the degree to which the systems are designed to respond to their needs and contain measures they find meaningful. When field staff found the measures relevant to their day-to-day work, they were more likely to be satisfied with the system. At the other end of the spectrum were cases in which field staff felt that the data aggregation process rendered the system's measures meaningless. In these cases, they saw little value and even potential harm in the system.

Chart 5: The utility of their agency-level measurement system to survey respondents (N=57)

Not useful; 18%

Very useful; 30%

Somewhat useful; 53%

Note: Survey respondents tended to be headquarters-based staff.

Staff charged with developing and managing agency-level measurement systems are sometimes more satisfied with the process outcomes than they are with the actual data. These staff members highlight how the discussions to identify meaningful organizational-level measures have helped staff de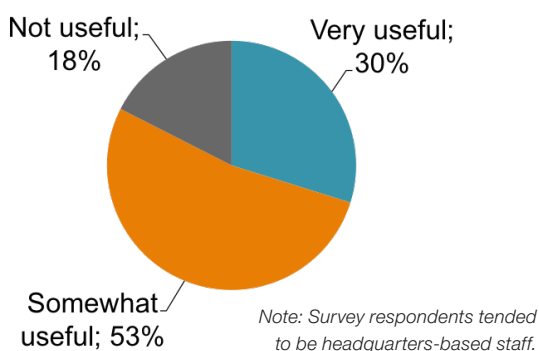velop a shared understanding of the organization's goals and the interrelatedness of their work. They are pleased with staff's willingness to reflect on work and improve. They acknowledge that efforts to measure results at the agency level revealed weaknesses and gaps in general

> The consultation process has prompted a new forum for staff to share experiences across projects and countries and for renewed engagement and centralizing around organizational objectives.
>
> *– NGO staff member*

M&E capacity, and drew needed attention and investment to strengthening that capacity. In some cases, they note a cultural shift and capacity building around valuing measurement and results. They also appreciate the development of a repository for general information about their organizations' work.

Can one system meet everyone's needs?

Many organizations have found it difficult to develop an agency-level measurement system that meets the needs of all staff. As previously noted, in indicator-based systems, when project-level indicators are not standardized, they are still meaningful at the project level, but they are very difficult to aggregate and make meaningful at the agency level. This problem is exacerbated in multisector NGOs. This approach can also entail managing a list of 300-500 indicators: a measurement and analysis burden that can be impossible to sustain.

Some indicators are easily standardized, while still maintaining project-level utility, thus satisfying both local project staff and leadership. Other indicators are harder to standardize in a meaningful way, so attempts to

> The more data are aggregated, the less useful they actually are. Meaningful indicators are found at project level.
>
> *– NGO staff member*
>
> We have long captured a large list of "shared indicators" that are mainly program outputs – number of children receiving vitamin A supplementation, number of people trained in various topics, and so on. While these are of interest in terms of speaking to a public about our reach, they do not necessarily provide any strategic or qualitative information.
>
> *– NGO staff member*

standardize them may cause them to lose value for country-level staff and render them difficult to interpret even at a headquarters level.

In alternate systems, such as those that rely on meta-evaluations or samples of existing evaluations, selecting themes or samples that are useful beyond the sectors from which they are picked is challenging. In these cases, the agency might learn more from the process about evaluation capacities and gaps than about project results.

Taking into account an organization's structure is important for getting buy-in into an agency-level measurement system. The more centralized an organization, the easier it is to agree on a system design, standardize measures, and gather and manage useful data. While this finding does not dictate a particular structure, it suggests that a less-

centralized organization will need to invest more time and staff hours to get buy-in, build, and manage a system. It also indicates that a system needs to reflect an organization's structure. For example, if an organization is partner-driven, then the system design process should be primarily bottom-up.

It is challenging, though feasible, to produce a measurement system that produces data that are meaningful to country staff and program managers, as well as to senior management and boards. This is most successfully done when an organization's global strategic plan and country plans are closely aligned, allowing for linked metrics. It is also easier to do when an organization is engaged in fewer program areas, although health-focused organizations will note that, even with fewer sectors, much diversity remains.

In another approach, some NGOs design systems that include both a menu of optional standardized indicators that help simplify program staff's indicator development work and a smaller set of core standardized indicators that everyone must measure for agency-level reporting purposes. In this case, program staff benefit from access to already standardized and defined indicators, making the burden of reporting on some of them lighter. With this approach, organizations must be careful to ensure that the data from the indicators in the menu are not aggregated, since optional reporting on them will provide an incomplete picture of the organization's accomplishments. While basket indicators allow greater specificity for sector and project-level indicators, they can entail a difficult analysis burden for system managers.

What information is used and how?

Information is most often used for accountability to leadership and the board. Of the submitted case studies involving systems already partly or fully in use, three-quarters reported using the data for that purpose. Two-thirds reported using the data for communications, project management, and ensuring technical quality. The information was less likely to be used for strategic decision making, learning, and business or proposal development, although there were cases in which this happened. In all of these cases, organizations noted that the measures help raise important questions that require further exploration, but they do not provide answers regarding what works or how.

Misuse or failure to use information

There are many reasons why data from agency-level measurement systems are not well used. Organizations most often explain that the information did not respond to strategic decision-making needs, the aggregated information was not meaningful, there were

Those are good numbers.
Don't just throw them away.

freshspectrum.com

Chart 6: Uses of agency-level measurement system information (N=17)
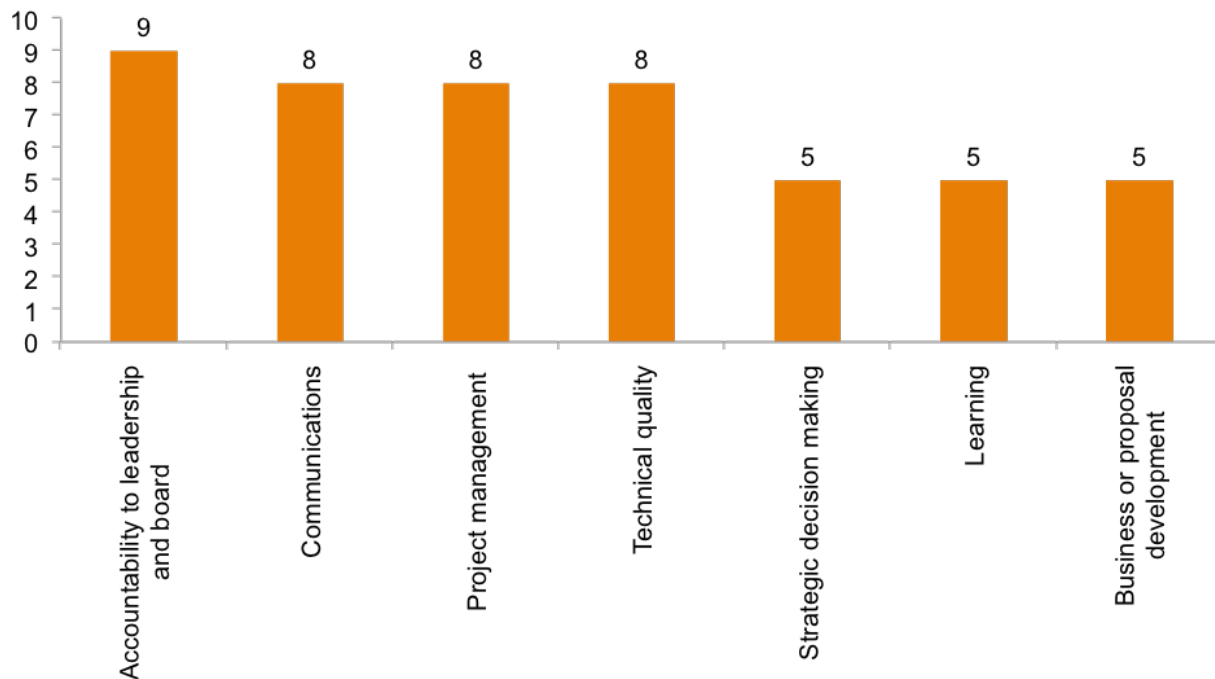
problems with data quality, the information was not timely or relevant, and/or the information was difficult to access because of the data management platform in use.

Organizational priorities and staff skills often do not foster data use. Staff report a lack of a culture promoting using data to inform decision making, and a lack of time for or prioritization of data use. Staff members also often lack data analysis and interpretation skills. They also often note a rewards structure that values bringing in money over accomplishments.

With agency-level measurement data, there is also a risk of misuse. Project-level data might get used for comparisons across countries when differences in those country contexts make those comparisons inappropriate. Performance information provided for learning purposes to inform improvements could be used to inform future investments. Or information intended for internal learning could be used externally, creating a reputation risk. These risks point to the need to constantly communicate the appropriate uses of the data the system produces and to have someone positioned to enforce that usage.

> Higher-level executives and board members assume this system will provide a bottom-line indicator of impact and effectiveness (as profit does in the business world) and seem annoyed when the system can't produce the results expected.
>
> *– NGO staff member*

C. Costs and benefits

The most frequently mentioned benefit is that the process of developing the system helped the NGO understand and address its progress and results, project management and M&E capacities, and the capacities of its information management systems. These new understandings, shared among staff, leadership, and boards, have sometimes led to new investments or significant adjustments. Staff talk about the evolution of an organizational culture around measuring and understanding effectiveness in data-driven, strategic, globally aligned ways. They also talk about how these measurement systems and the process of developing them can be used to break down silos and unify an organization or a (con)federation by helping staff see how their programs fit together and contribute to a broader goal.

The biggest cost may be the cost of setting up and implementing an information management system, as described above. There are also clearly staff costs, whether an organization hires new staff to help implement the system or dedicates existing staff to the task. Finally, there are opportunity costs associated with using staff time for this purpose when the time could be used to help improve basic program and project-level measurement.

Most headquarters-based evaluation staff felt that developing an agency-level measurement system was ultimately a worthwhile endeavor. All case study contributors agreed, and only 15% of survey respondents felt the costs outweighed the benefits. Most, reflecting on their systems' purposes, users, and the costs, felt that their

systems required adjustments, but would not recommend getting rid of the systems entirely. However, most case study writers acknowledged that field staff carry the majority of the data collection burden and sometimes see little benefit. In one case in a partner-driven organization with a system that serves the needs of leadership, field staff felt the system was not worthwhile.

D. Lessons learned

Input from participants with established agency-level measurement systems offer a number of lessons for others thinking about whether and how to design such systems.

1. Agency-level measurement systems can be difficult to set up in very decentralized organizations where decision making depends on reaching consensus among many different parts

> You learn things about yourself that you wouldn't otherwise know, or you wouldn't have the hard numbers to back up the conversation to take the emotion out of it. When you look at numbers collectively, reflecting on all of us, it facilitates the conversation about what we do well and what we don't do so well.
>
> *— NGO staff member*

(whether a confederation or an international NGO that has devolved significant power to country offices), since these systems require reaching agreement on a system's purpose and contents and probably also require standardizing measures.

2. Systems designed during periods of organizational change poses unique

---

Chart 7: Challenges faced by agency-level measurement systems (N=54)

Challenge	Percentage
The complexity, time, and cost required is higher than expected.	63%
Country staff find the system too burdensome.	46%
People do not trust the data.	32%
Executives want information quickly.	32%
The data provided are not actionable.	26%
The information the system produces is seldom used.	24%
Executives are disappointed with what the system provides.	20%
The system does not have a real owner.	19%
The complexity, time, and cost required outweighs the benefits yielded.	15%
We have/do not face(d) any challenges with our system.	7%
The system works well, but the organization has lost interest in using the information produced.	6%

Note: Survey respondents could check all that applied.

challenges (and opportunities) and before there is clarity about the direction of that change. However, once there is a clear vision for the outcome of the organizational change, setting up an agency-level measurement system can reinforce that new vision. However, even with a clear vision for a changing organization, it can be difficult to launch an agency-level measurement system, since during that time, country offices must implement many changes simultaneously, while still running their programs. At the same time, large change processes often include significant staff turnover, meaning that those introducing new systems or practices have to do so repeatedly, as new staff come on board.

Further, introducing an agency-level measurement system is an organizational change process in its own right, with all the potential for staff questions, anxiety, and risk of resentment that implies. As such, it requires clear communication and behavior modeling from senior leadership.

3. There can be a lag time of a few years between when an agency-level strategy or priorities are set and when programming begins to reflect that vision. In the meantime, it might not be clear what needs to be or can be measured. This may require taking a gradual approach to building an agency-level measurement system, starting with basic project data, then sectors and what is already being measured, and later strategy indicators, as the strategy becomes clearer.

4. Developing an agency-level measurement system often takes more time, funding, and skills than anticipated. The experiences of various organizations point to a five-year period for system design, development, and implementation. One respondent warns that by the time a system is producing data, an organization's priorities might have shifted, making it no longer relevantly focused. Avoiding cost overruns requires getting clarity on information system needs prior to investing in it, which may be an unaffordable luxury if data are coming in and need to be managed. Poor data quality can hinder data use; yet the necessary capacity-building investments require time, funding, and staff, and may take time to produce visible results.

> One NGO doubled the size of the team dedicated to managing its agency-level measurement system once it better understood the system's demands. Another went from having one M&E staff member working part-time on its system to a team of two full-time and two part-time staff members dedicated to it. In both cases, the original managers were M&E staff, while the expanded teams also included project managers and IT staff.

5. Managing expectations depends on having a realistic sense of what the chosen system design demands and can deliver, as well as the resources the organization can dedicate to it. In successful cases, NGO staff and leadership have come to agreement on the best possible agency-level measurement system design given the resources available.

6. Organizations need to dedicate sufficient numbers of skilled M&E staff members to build and maintain a useful system. The required skills include data collection, management (with a user-friendly

platform), analysis, and presentation. The M&E staffing should include a small, dedicated team in charge of building and managing the system, as well as time allocated by M&E staff at all data collection, management, and analysis points in the organization.

7. While M&E capacity is paramount to launching and implementing a useful agency-level measurement system, most NGOs recognize that they also need a **user-friendly platform where staff can easily store, access, and analyze the data collected, and use the data for their own purposes.** Where these systems have been successful, M&E and IT staff can effectively communicate with each other and have collaborated together well. They understand each other's capacities and constraints, as well as those of their program colleagues, both

> The bulk of the work of populating the system falls to field teams who are very busy and struggle with connectivity problems. At the same time, there doesn't seem to be a real pay-off for them with actionable or insightful data.
>
> *– NGO staff member*
>
> As much consultation as there has been throughout the development of the indicators with field teams, it has been really difficult to justify all the time and cost invested in the system, when there are so many pressing needs at the country-level for more technical support and boots on the ground.
>
> *– NGO staff member*

in headquarters and country offices. They recognize that the measurement design drives the system, and that the platform is there to support it. These platforms have ranged from Excel workbooks to off-the-shelf systems to customized systems. In all successful cases, organizations have understood that the human interaction around the platform is more important than the system itself.

8. If a system only requires data collection on an annual or six-month basis and is not well integrated into other processes, staff may always see it as something added on to their daily tasks. In addition, executive leaders may have the unrealistic expectation that measurement happens more frequently. More frequent measurement and making measurement part of staff's job descriptions could get measurement into staff's DNA. But accomplishing this is difficult, given limited resources and high demands on staff time.

9. Agency-level measurement systems often do not respond to country program needs. The aggregated data are often too general to inform country-level strategic decision making. This reduces field staff buy-in and leaves them seeing the systems as an added burden with no value-added for them. This is especially challenging, since much of the data the system requires must come from them. Ensuring value for field staff requires additional investment of time in the design stage: identifying measures beneficial for multiple user groups; and helping field staff build their data collection, management, analysis, and reporting skills, in ways that support both the agency-level measurement system and their own daily work.

IV. Recommendations

Based on the findings of this study, a number of recommendations emerge for those developing or considering developing an agency-level measurement system.

1. Ensure a clearly defined purpose and users. Invest sufficient time and resources in defining and agreeing on the purpose, users, and content of the system, as well as reviewing this over time. Make clear choices, ensuring all involved understand that no one system can satisfy the data demands of all stakeholders.

2. Align the agency-level measurement system with agency values, mission, and strategy. The closest alignment should be with strategy and organizational theory of change, since those describe an organization's medium-term objectives and how it hopes to achieve them. Make sure your system does not fall between the cracks between what senior management wants to know and what country program staff want to know, since then it will not be used by either.

3. Foster an organizational culture of evidence-based decision making. Organizational culture refers to the set of often unspoken values, belief systems, and assumptions – widely shared among staff – around what matters in an organization. In an evidence-based decision-making culture, people are hired and rewarded (formally and informally) in part for effective use of data,

including agency-level data. Staff who use data are held up in such organizations as heroes. Executive leaders align their decision making to such data. And the symbols and language that the agency uses reflect the importance of such data. Achieving this might entail a cultural shift; this requires change management support and typically takes seven to 10 years to fully accomplish, although an organization may begin to see change sooner.[7]

4. Invest in cultivating ongoing ownership and championing of the agency-level measurement system at the top, including dedication of executive leadership's time to understand, analyze, and use the information generated. This will spur broad buy-in, and will encourage adequate and ongoing investment in the system's development, management, and use.

5. Allocate sufficient resources (human resources, funding, time) to build and implement the system.

- *A team of both M&E and IT staff members needs to be dedicated to managing and coordinating the system.*

---

7 See Laia Griñó, Carlisle Levine, Stephen Porter and Gareth Roberts, *Embracing Evaluative Thinking for Better Outcomes: Four NGO Case Studies*, InterAction and CLEAR-Anglophone Africa, June 2014, http://www.interaction.org/node/3818. See also Edgar Schein, Organizational Culture and Leadership, Wiley and Sons, San Francisco, 2010.

- An NGO needs to prioritize **staff capacity building** for data collection and analysis, since this is essential for ensuring data quality and use. An assessment of existing monitoring capacity before developing a system can help steer investments and ensure realistic expectations. While the capacity is being built, the system should be rolled out slowly so as to not outpace the actual capacity levels.

- Leadership needs to understand that developing such a system takes **time**. It can take up to five years before a system begins to provide actionable data. This includes a year or two to design a system, given the requirements of consultation and getting staff buy-in, and three more years to fully roll it out and implement it, given capacity building needs.

  » Go slow and offer lots of technical assistance.

  » Start small, perhaps starting with simpler measures, such as inventories of projects and numbers of participants reached; then add more complicated measures or launch a pilot in a few countries before rolling the system out more broadly.

- Leadership also needs to be realistic about **costs** over time, as well as **opportunity costs,** and ensure that the benefits of investing in an agency-level measurement system outweigh the costs of not investing staff time elsewhere.

6. Include field expertise to design a system that is realistic and relevant to both top-level leadership and field staff. One staff member observed that those developing the systems (and the IT systems) seem to be more effective if they have extensive field experience.

7. Take steps to ensure broad staff buy-in into the system. This can be achieved in a variety of ways.

- **Hold collective discussions** on how metrics relate to strategy or priorities, why metrics are being measured, choices for metrics design, data collection, and management approaches. This can lead to more buy-in and a more meaningful and useful system. Building on existing systems or experiences collecting data at other levels (country, regional) within an organization can also foster buy-in.

- **Constantly communicate** the purpose of the system (especially if it is for learning), reporting requirements, and timelines with plenty of advanced notice. Communicate in all official national language(s) used by country staff.

- **Tie indicators to measurements relevant to country office activities or to the part of the organization that is responsible for collecting the data** so that the system provides information useful for regular decision making at all levels. This can be accomplished by helping country offices develop strategies and action plans that align with the organization's global strategy and action plan. It can also be done by basing the selected indicators on indicators country offices already collect within their projects. Indicators designed in this way can be seen by country teams as "time savers,"

> To build and manage an agency-level measurement system, an international NGO requires a high-level owner for the system to motivate buy-in and maintain momentum, at least two full-time monitoring and evaluation (M&E) staff members, and full-time information technology (IT) support. The M&E and IT staff must have a shared understanding of the system and must collaborate well together. This core team will also require support from staff members with sector-specific M&E capacities, as well as dedicated time from staff members who are charged with data collection.
>
> *– NGO staff member*

since if the organization did not offer them these indicators, they would have to create them for their projects.

- ***Make data collection, management, and analysis as simple as possible.*** Provide guidance, tools, and technical assistance. Having a system that is both online and offline maximizes access to data and facilitates data management. Do not create parallel measurement and data inputting systems for the field, since that creates unnecessary work burden, resentment, and non-use.

- ***Give staff the needed resources (time, tools) to increase their buy-in.*** Investment must be made in staff's data collection, management, analysis, and interpretation capacities.

- ***Provide all staff with access to the data.*** Having an online and offline feature facilitates this. Showing all involved in data collection how the data are used and making the data useful to the data providers can further buy-in.

- ***Make measurement and monitoring part of job descriptions, workplans, and performance evaluations*** in order to make staff accountable for data collection, management, analysis, and use.

- ***Integrate the agency-level measurement system into planning and management processes.*** Use it to ask about performance in certain areas and to determine what needs to be changed and what requires more investment.

8. Indicators must be standardized for aggregation. Aggregating non-standardized measures is too complex and takes too much time. In certain sectors, some indicators have already been standardized. Where they have not and where it is possible to standardize measures, an organization needs to invest the time and staff skills in doing so; this can be a sizeable investment. For indicators that do not lend themselves to standardization, consider whether it would be better to use basket indicators or simply leave those indicators out of the system.

9. Recognize that indicators do not tell a full story. They raise questions. Understanding context is important. Mini-case studies and evaluations can provide depth, while indicators show breadth.

10. Ensure that the information management system is user friendly. It should make data accessible for all staff. Inputting and using the data should be as easy as possible, offline as well as online, and should accommodate for low bandwidth. An IT specialist should be available to provide ongoing technical assistance and make adjustments to the system as needed.

Appendices

Authors

Carlisle Levine

Carlisle Levine is an international development, peacebuilding, and advocacy evaluator with 24 years of international development experience and 17 years of evaluation experience. She is a skilled facilitator and trainer with expertise in leading complex evaluations, building staff evaluation capacity, developing monitoring and evaluation systems, fostering collaborative learning processes within dispersed and diverse teams, and developing knowledge sharing systems. Her work has contributed to organizational strategic decision making and more effective practices. Prior to launching her own business, Carlisle worked for CARE USA, Catholic Relief Services (CRS), USAID, and the Inter-American Foundation and with InterAction; her work has taken her to Latin America, Central and West Africa, and Europe.

While working for CRS, Carlisle led an agency-wide process to develop common indicators for CRS programs for the purposes of measuring agency-level results and facilitating program design. Carlisle took a field-based, consultative approach that sought to ensure broad utility and ownership. She also built off of the experiences of other international NGOs further advanced in this process. During her tenure at CRS, Carlisle produced a concept paper which had broad buy-in, and two sectors developed and implemented measures. Based on this experience, valuable lessons were learned.

While working at CARE, Carlisle was peripherally involved in the revision of one agency-level program result measurement system and the development of another that was organization-wide. As a consultant, Carlisle has worked with a number of clients, such as The MasterCard Foundation, the Boulder Institute of Microfinance, Bread for the World, Rise Up, and the Rights and Resources Initiative, to develop agency and program-level results measurements, gaining broad experience and learning important lessons in the process. While serving as InterAction's Senior Evaluation Advisor (consultant), Carlisle worked closely with InterAction to develop papers and lead workshops for members related to evaluative thinking and local ownership in evaluation. These efforts, building on members' and others' experiences, produced practical guidance that InterAction members and others are using.

Tosca Bruno-van Vijfeijken

Tosca has worked in international development and organizational development, change processes, and effectiveness issues for over 25 years. She has focused on evaluation for the last

eight years, is an active member of the American Evaluation Association (AEA), and has undertaken several strategic, as well as meta-evaluations, for Plan International. As a consultant, Tosca has evaluated the organizational change processes of Save the Children, Oxfam, CARE, and Amnesty International. As a "pracadamic" based at the Maxwell School of Citizenship and Public Affairs at Syracuse University, she bridges the field of academic knowledge production with that of NGO practice through her role as director of the Transnational NGO Initiative. Tosca is a board member of InterAction, as well as a member of the advisory panel of Charity Navigator. She has in-depth experience with qualitative methods such as participant observation, participatory assessment, and focus group discussions, and has conducted leadership-level interviews with hundreds of international NGO leaders over the past 12 years.

Sherine Jayawickrama

Sherine Jayawickrama is an independent consultant working with international NGOs and foundations on efforts to enhance organizational governance, adapt institutional architecture, improve organizational learning, build trust and communication at the CEO level, and evaluate organizational evolution and development processes. She has 19 years of experience focused on international development from various angles, including from the vantage point of a major international NGO, a USAID-funded project in Sri Lanka, and a leading U.S. university.

From 2009 to 2013, Sherine managed the Humanitarian & Development NGOs domain of practice at the Hauser Center for Nonprofit Organizations at Harvard University, where she managed and wrote for a blog on international NGOs, and served as executive director of the NGO Leaders Forum, an annual retreat for the CEOs of America's major international NGOs, where she regularly produced briefing papers for a CEO audience. Since July 2013, she has been an independent consultant, while still playing a coordination role for the NGO Leaders Forum and maintaining an association with Harvard University.

From 1999 to 2008, Sherine worked at CARE USA in a variety of positions, including deputy regional director for Asia, senior policy analyst and special assistant to the president. Earlier in her career, Sherine worked on freedom of expression issues at PEN American Center in New York City, and with environmental regulators and private sector leaders on environmental policy issues in Sri Lanka at the USAID-funded Natural Resources & Environmental Policy Project (NAREPP).

Methods

The consulting team utilized six methods to gather information from different perspectives and in different ways. These methods included the following, each of which is described in more detail below:

1. **World café** – The consultants led a world café with InterAction member CEOs who were participating in the annual InterAction CEO retreat.

2. **Focus group discussion** – The consultants led a focus group discussion with evaluation staff from InterAction member international NGOs who were participating in the annual InterAction Evaluation and Program Effectiveness Working Group roundtable.

3. **Interviews** – The consultants conducted interviews with 26 people representing current and former InterAction member NGO staff members and consultants.

4. **Survey** – The consultants shared a survey with InterAction members and more broadly via the listservs Pelican and MandENews, which attract evaluators and other international NGO staff members, among others living and working globally.

5. **Case studies** – Seventeen InterAction member international NGOs contributed mini-case studies as input to this study.

6. **Literature review** – The consultants also read presentation materials developed by InterAction member organizations, materials developed by InterAction's Evaluation and Program Effectiveness Working Group, and documents written by consultants. InterAction's website contains the first two groups of documents, while the bibliography lists the last group.

Instruments

CEO WORLD CAFÉ

Focus

Facilitated CEO discussion of agency-level measurement: what are our expectations as top leaders, and what have been our experiences?

Session details

InterAction CEO retreat, December 3, 2015, 3:45-4:45pm

Session description

Some international NGOs have made significant efforts to put in place agency-level measurement systems as part of a concerted focus on better accountability for results, as well as learning for improvement. What have been your drivers for such undertakings? What have been your assumptions, hopes, and expectations for these systems as CEOs? And what have these systems actually delivered thus far for you? The interactive world café style group discussion was structured to allow participating CEOs to discuss the advantages, limitations, risks, and tradeoffs involved in attempting to measure results at the agency level.

Questions posed

Question 1: What have been your drivers, assumptions, and expectations for agency-level measurement?

Question 2: What have these systems actually delivered thus far for you?

Question 3: If you were to advise leaders on developing and implementing an agency-level measurement system, what would you tell them?

Design

- Participants: 50-60 CEOs.
- Tables of 5-6 per table; 10 tables total.
- Each table has flip chart or at least flip paper.
- Each group has one person who moderates and one who takes notes on flip chart.
- Five-minute introduction of the topic and consultants.

- Three questions; people rotate after 15 minutes to next table for a total of 30 minutes, with each person responding to two of the three questions; 20 minutes for reporting back (1-2 observations per table, popcorn style, keep building to additive list).

EPEWG FOCUS GROUP DISCUSSION

Focus

InterAction members' experiences with and perspectives on agency-level measurement.

Session details

Focus group discussion with the InterAction Evaluation and Program Effectiveness Working Group, October 15, 2015

Session description

Facilitated discussion with working group members concerning agency-level measurement.

Questions

1. Number of working group member NGOs interested in developing, in the process of developing or with agency-level measurement systems? (show of hands)

2. What is the primary reason that NGOs want to develop agency-level measurement systems?

3. What are some of the forms that your organizations' agency-level measurement systems (or equivalents) take? What are their key building blocks – project data, something else?

4. Who is the primary audience for your agency-level measurement data?

5. How is (or could) data generated by your organizations' agency-level measurement systems be used and by whom?

6. What are some of the challenges your organizations have faced or might face in developing and using agency-level measurement systems?

7. What have been some of the benefits your organizations have experienced or could experience in developing and using agency-level measurement systems?

8. What choices have your organizations made or could your organization make to ensure that the benefits of having agency-level measurement systems outweigh the costs of developing and managing them?

INTERVIEW PROTOCOL

The consultants adjusted the interview protocol used per interviewee. Below is an example of a protocol used with a case study participant. The other protocols were variations on this one.

Measuring International NGO Agency-Level Results: Developing a White Paper
Case Study Participants: Interview Protocol
January 2016

Project overview

For many years, both donors and international NGOs have sought to effectively assess the sum total of their work and tell their organizations' stories. However, the complexity of international assistance processes and international NGOs' work have made capturing agency-level results challenging.

Recently, a number of InterAction members have come together to learn from each other's experiences. These efforts represent diverse approaches to agency-level results measurement with varying degrees of success and setbacks. Together, they can offer valuable guidance for the broader field. This white paper seeks to contribute to that learning, by taking both a broader and deeper look at InterAction members' experiences, and also drawing in lessons learned from outside.

1. **Motivations:**

    a. What were your organization's motivations to develop an agency-level measurement system?

    b. What were your organization's expectations, as well as assumptions, for what this system could accomplish?

    c. Was there broad agreement on these points within your organization, or were there differing points of view? Please explain.

2. **Assumptions:** What were your organization's assumptions regarding the capacities and other elements required to develop, manage, and use an agency-level measurement system?

3. **The Agency-Level Measurement System:**

    a. Please describe the specific nature of the agency-level measurement system you have put in place (or are developing).

    b. Please explain why your organization designed or is designing your system in this way.

4. **Use:**

   a. When designing your agency-level measurement system, who were the primary and secondary intended users, and how was it anticipated that they would use the data?

   b. Please describe how, if at all, the data generated by your organization's agency-level measurement system are actually being used or will be used and by whom.

   c. Is there anyone who you thought would use the data, but is not? If so, please explain.

   d. If there is a difference between intended and actual users, please explain.

5. **Satisfaction:**

   a. How satisfied is your organization with the system your organization adopted?

   b. Is there variation in levels of satisfaction among different staff members? Please explain.

   c. Are there parts of the system that are more satisfying? Please explain.

   d. Are there parts of the system that are less satisfying? Please explain.

6. **Challenges and Lessons Learned:**

   a. What challenges has your organization faced in developing, implementing, and using data from your agency-level measurement system?

   b. How has your organization responded to these challenges?

   c. What lessons emerge from responding to these challenges that could be valuable to other organizations?

7. **Costs and Benefits:**

   a. What have been the benefits of your agency-level measurement system?

   b. What have been the costs of developing and managing it, including actual and opportunity costs?

   c. Have the benefits outweighed the costs, or vice-versa? Please explain.

   d. Based on your responses above, if you could do it over again, would you design your system in the same way? Or are there changes you would make? Please explain.

SURVEY

The following is the text of the survey used for this project.

Agency-level measurement systems seek to effectively assess the sum total of an organization's work and tell the organization's story. This is a challenging endeavor for international NGOs (INGOs) because of how varied their work is (in terms of sectors, geographies, donors). A number of InterAction members have come together to learn from each other's experiences, both successes and setbacks. This survey expands that learning effort by seeking the experience of a broader group of INGOs. Data generated by this survey will facilitate the development of a white paper on agency-level measurement that aims to be a resource to INGOs in the future. We will be happy to make the final product available to you upon completion.

Your candid responses will be helpful input into the white paper. The study team will keep your responses confidential, and you may choose to respond anonymously. Thank you for taking the time to complete the survey!

Overview

1. Does your organization have an agency-level measurement system? Select one response.
   - ❏ Yes, it is well established.
   - ❏ Yes, and it is in the process of being modified.
   - ❏ Yes, but it will soon be replaced by a new system.
   - ❏ No, but we had such a system in the past.
   - ❏ No, but we are currently developing such a system.
   - ❏ No, but we are interested in developing such a system in the future.
   - ❏ No, we have not considered adopting such a system yet.
   - ❏ Other (please specify).

2. What approaches does/will/did your organization's agency-level measurement system feature? Select all that apply.
   - ❏ Standard indicators that projects are required to use.
   - ❏ A menu of common indicators from which projects can select.
   - ❏ Common indicators that map to higher level indicators aligned with strategy, mission or vision.
   - ❏ Meta-reviews or meta-evaluations by program area or sector.
   - ❏ Other (please specify).

Motivations and Expectations

3. What was/is your organization's motivation for developing an agency-level measurement system? Please select all that apply.
   - ❏ Our board mandated that we develop such a system.
   - ❏ Our senior leadership mandated that we develop such a system.

❏ We wanted to enhance our learning about our results at an agency level.
❏ We wanted to enhance our accountability to donors.
❏ We wanted to enhance our accountability to program (or intervention) participants, and the communities in which we work.
❏ We wanted to have a better sense of how many people we are reaching with our programming and/or policy work.
❏ We wanted to be better able to market what our organization achieves.
❏ Other (please specify).

4. If your organization's motivation for developing an agency-level measurement system has changed since you started implementing it, please explain.

5. Who were/are the main intended audiences for the information generated by your organization's agency-level measurement system? (If your organization has had multiple systems, please refer to the most recent one.) Select all that apply.
❏ Board of directors
❏ Senior leadership
❏ Marketing and communications staff
❏ Fundraising staff
❏ Programming staff
❏ Policy and advocacy staff
❏ Donors (institutional and individual)
❏ National governments in countries where programs operate
❏ U.S. government
❏ General public
❏ Other (please specify)

Utilization
6. How useful is your organization's agency-level measurement system to you? Please select from the drop-down menu.

7. Please share what your organization has done to promote the use of data collected by the agency-level measurement system.

8. If your organization's agency-level measurement system has been useful, please describe how it has been used and by whom. If your organization's system has not been useful, please explain why that is/was the case.

Challenges and Lessons
9. What are the major challenges that your organization's agency-level measurement system has faced (or faces currently)? Please select all that apply and provide further detail in the text box below.

❏ The complexity, time, and cost required to build and maintain the system was a lot more than anticipated.
❏ The complexity, time, and cost required to build and maintain the system outweighs the benefits yielded.
❏ Country or field leaders and staff find the system too burdensome and do not see value returning to their work.
❏ People do not trust the data that comes from the system due to problems with data quality.
❏ Executives do not see usable information quickly enough.
❏ Agency-level information that the system produces does not have a real owner.
❏ Agency-level information that the system produces is seldom used.
❏ Executives have realized that the system cannot provide what they expected.
❏ The system works well, but the organization has lost interest in using the information produced.
❏ The data provided by the system is not actionable.
❏ We have/do not face(d) any challenges with our system.

Please share more insight into the challenges, if any, faced by your organization's agency-level measurement system.

10. If you have observed any unintended consequences (positive or negative) of having an agency-based measurement system, please explain.

11. Based on lessons learned by your organization, what advice would you give an organization that is considering options for effectively measuring results at the agency level? Please be as candid and creative as possible!

Organizational Information
12. In how many countries does your organization work?

13. In how many sectors does your organization work?

14. Approximately how many M&E staff does your organization employ at various levels?
   • Headquarters/funding office level
   • Regional level
   • Country level

15. What are your organization's primary sources of funding? Select all that apply.
❏ U.S. government and/or other bilateral or multilateral donor funding
❏ Individual donations (small sum donors, high net worth donors, etc.)
❏ Foundation or corporate funding
❏ A mix of all of the above
❏ Other (please specify)

16. What is your organization's annual budget?

17. Is your organization a member of InterAction?

18. What is your position within your organization?
   ❏ M&E (or related) staff
   ❏ Program staff
   ❏ Policy staff
   ❏ Marketing and communications staff
   ❏ Senior leadership
   ❏ Other (please specify)

19. Where are you located?
   ❏ Headquarters or funding office
   ❏ National member or affiliate office
   ❏ Regional office
   ❏ Country office
   ❏ Other (please specify)

20. (Optional) Name of organization

21. (Optional) Name and email address for follow-up.
   • Name
   • Email address

List of case studies

The following InterAction members contributed case studies as input to the development of this white paper:

1. ACDI/VOCA
2. American Jewish Joint Distribution Committee
3. Bread for the World
4. CARE International
5. Catholic Relief Services
6. ChildFund International
7. Heifer International
8. Helen Keller International
9. International Rescue Committee
10. Jhpiego
11. Lutheran World Relief
12. Mennonite Central Committee
13. Mercy Corps
14. Management Sciences for Health
15. Save the Children
16. World Learning
17. World Vision

Interviewees

1. Paul Amendola, Technical Advisory, Health Information, Health Unit, International Rescue Committee
2. Yeva Avakyan, Senior Gender & Evaluation Advisor, World Vision US
3. Yolanda Barbera, Deputy Director for Strategic Programs, Health Unit, International Rescue Committee
4. Chip Barnett, former Senior Technical Advisor, Organizational Measurement, International Rescue Committee
5. Ken Berger, former President, Charity Navigator, and Managing Director, Algorhythm
6. Muluemebet Chekol Hunegnaw, Senior Director, Monitoring & Evaluation and Knowledge Management, Save the Children US
7. Mark Anthony R. Dasco, Program Director, Philippines National Office, ChildFund Philippines
8. Zaira Drammis, International Head of Monitoring and Evaluation, ActionAid International
9. Ximena Echeverria M., Project Manager, Project/Program Information and Impact Reporting System, CARE International Secretariat
10. Matthew Forti, Managing Director, One Acre Fund USA and former Head of Performance Measurement at Bridgespan Group
11. Kent Glenzer, Dean of the Graduate School of International Policy and Management and Associate Professor at the Monterey Institute of International Studies, and former head of Learning, Evaluation and Accountability (LEAD) at Oxfam America and former Director of Impact Measurement and Learning at CARE USA
12. Allen Grossman, Board Chair, Mercy Corps, and Senior Fellow, Harvard Business School
13. Jeanne Jantzi, Co-Area Director, Asia, Mennonite Central Committee
14. Neal Keny-Guyer, CEO, Mercy Corps
15. Maliha Khan, Independent Consultant, former Director of Learning, Evaluation and Accountability at Oxfam America and former Director of Program Impact at CARE USA
16. Asma Lateef, Director, Bread for the World Institute
17. Adeeb Mahmud, Director, FSG
18. Velina Petrova, Director, Program Strategies, Assessment & Learning , ChildFund International
19. Emily Sinn, Program Manager, Tola, Mercy Corps
20. Glenn Strachan, Chief Information and Technology Officer, Jhpiego
21. Holta Trandafili, Design, Monitoring and Evaluation Specialist, World Vision US
22. Bernard Vicary, Independent Consultant and former Director, Program Management Information Systems, World Vision International
23. Barbara Willett, Monitoring and Evaluation Lead, American Red Cross, and former Senior Technical Advisor, Design, Monitoring and Evaluation, Mercy Corps
24. Jeff Yaschik, Director, Financial Planning, Reporting and Analysis, CARE USA
25. Anna Young, Senior Director, Strategy and Learning, Mercy Corps
26. Kristen Zimmerman, Learning and Evaluation Coordinator, Mennonite Central Committee

Bibliography

In addition to reviewing a range of international NGOs' in-house working documents on agency-level measurement and the individual case studies of 17 InterAction member organizations, the consultants also reviewed the following readings:

InterAction *Evaluation Working Group presentations on agency level measurement*, https://www.interaction.org/work/monitoring-evaluation.

Kara, Nabihah and Matt Forti. *Making Measurement Work in Large, Complex Organizations*, Bridgespan Group blog post, January 16, 2014. http://www.bridgespan.org/Blogs/Measuring-to-Improve.aspx#.VwLepE8rLct.

Schein, Edgar. *Organizational Culture and Leadership*. San Francisco: Wiley and Sons, 2010.

Simister, Nigel. *Summarising Portfolio Change: Results Frameworks at the Organisational Level*, M&E Paper 10, INTRAC, January 2016. http://intrac.org/data/files/resources/888/Summarising-portfolio-change-results-frameworks-at-organisational-level-FINAL.pdf?utm_source=New+publications&utm_campaign=publications+january+2016&utm_medium=email.

Simister, Nigel. *Striking the Right Balance for Monitoring and Evaluation in Complex Organizations*, Evaluation/Learning Brief, No. 10, Danish Refugee Council, January 2016. https://drc.dk/media/2062363/e_l_brief10-striking-the-right-balance-for-monitoring-and-evaluation-in-a-complex-organisation.pdf.

Villanger, Espen, Rob Lloyd, Derek Poate & Merima Ali. "Why can we not demonstrate the difference that Norwegian aid makes? An evaluation of the results measurement system and practice," *Journal of Development Effectiveness*, Volume 8, Issue 2, 2015, p. 171-195. http://www.oecd.org/derec/norway/Can-We-Demonstrate-the-Difference-that-Norwegian-Aid-Makes.p

CRS CATHOLIC RELIEF SERVICES

PATH

World Vision®

Save the Children®

ChildFund International®

JDC

INTERNATIONAL RESCUE COMMITTEE

Lutheran World Relief

msh Management Sciences for Health

innovating to save lives
Jhpiego®
an affiliate of Johns Hopkins University

MERCY CORPS