

**Asia Regional Evaluation Evidence Exchange
October 19, 2015**

Setting the Stage: Failure, the Forgotten Twin of “Success” in Evaluation

Speaker: Stephen Mills, Technical Director, FHI360

Summary: Failure is the twin of success, yet we rarely talk about “the F word.” Most things in the private sector fail. Seventy-five percent of restaurants fail. Nine out of 10 ideas supported by venture capitalists fail. But there are no statistics in the public sector about failure; our language avoids admitting failure. Engineers without Borders and Fail Forward lead the way in addressing the issue of failure more openly and proactively.

The challenge for M&E specialists is that our findings rarely lead to good ideas being scaled up. We are working in silos that needs to be broken down. If we want to get our messages to benefit policies and programs, we need to partner with specialists in areas like marketing, advocacy and quality improvement. Our fields are complex and M&E is just one piece of the puzzle; our job is to make things simple for other people in the system. Consider the fact that the U.N. Sustainable Development Goals are really an evaluation exercise. They are meant to monitor how well our M&E tools work and how well we can simplify what we do.

Highlights:

- Evaluation findings need to be transferred into programs and policies by partnering with actors outside of the M&E field.
- M&E specialists need to take the complexity and challenges of the M&E tools and make them simple and easy for their partners.

Reflection: In the private sector, success and failure can be measured by whether something makes a profit or not. But in the public sector there are too many definitions of success; if a project fails, it can still continue. One of the strategies to help identify failure early in the project is to put an appropriate monitoring system in place. If the system captures the failure, it’s the action that happens after the failure is identified that matters. We need to ask ourselves how we can partner with other people to make sure our findings have an impact.

**Asia Regional Evaluation Evidence Exchange
October 19, 2015**

SESSION 1: QUALITY IMPROVEMENT IN EVALUATION - CARAVAN SESSIONS

1. Development Quality Statements of Work (SOWs) and Terms of Reference (TORs)

Station Manager: Suzanne Polak, Regional Learning, Monitoring and Evaluation Advisor, USAID/RDMA

Summary: USAID has developed a checklist on the USAID Learning Lab to help program managers design evaluation questions. The list serves the design stage of an evaluation. It helps plan an evaluation from the very beginning, including making an evaluation plan, managing evaluation, reporting and utilization. The checklist has three components:

1. Sources of questions – a list of available resources to help develop evaluation questions.
2. Prioritization – a list of questions to help determine which questions are most important to include in the evaluation.
3. Tips for writing good evaluation questions – well-written questions are essential to a successful evaluation.

Highlights:

- A checklist is not enough, it's a start. Critical thinking, dialog, creative process and flexibility in customization are more essential than just a checklist.
- A checklist is a mentoring and learning tool. It helps to formulate the right question.
- It's important to have a dialog about the questions.

Reflection:

- Checklists and guidelines are helpful if they are detailed and realistic. In evaluation, the question is how to use the checklist to ensure the SOWs and TORs are of the quality needed? The checklist in evaluation is a guide that helps the program team define the most appropriate evaluation question, what should be asked and when.
- Checklists are useful, but their usefulness depends on the context. For example, a checklist for an environment or human rights program won't work for gender issues. The methodology would be entirely different. Checklists can be used as a learning tool and a start in evaluation design, but should leave room for customization and flexibility, which later triggers critical thinking, dialog and creativity.
- It is important to maintain conversation between the technical office, M&E and program office in order to avoid problems.

**Asia Regional Evaluation Evidence Exchange
October 19, 2015**

SESSION 1: QUALITY IMPROVEMENT IN EVALUATION - CARAVAN SESSIONS

2. Managing the Evaluation and Team Composition

Station Manager: Mami Umayahara, Program Cycle Management Specialist, UNESCO Bangkok, Asia and Pacific Regional Bureau for Education

Summary: When we think about evaluation management, we come across 6 important steps:

1. Engaging stakeholders from the beginning, which includes managing people and beneficiaries.
2. Defining and agreeing on the purpose of the evaluation.
3. Designing the evaluation, which includes TOR development and team recruitment.
4. Data collection and analysis, gathering and presenting credible evidence.
5. Drawing and justifying conclusions and recommendations that could be utilized.
6. Sharing and using evaluation results.

These steps also include identifying risks and suggesting solutions to manage such risks at different stages of evaluation management. These risks are related to the evaluation team, data, reporting and utilization of results.

Highlights:

- An evaluation specialist with sectoral/thematic expertise should lead the evaluation team.
- An ideal evaluation team brings diversity in terms of gender, geographic and cultural background. There should be a clear division of labor.
- Stakeholder inclusion in the evaluation process increases ownership. Stakeholder communication during the evaluation process eliminates the negative effects on beneficiaries.

Reflection:

- A specialist that leads an evaluation team should have expertise in a particular area. In addition to solid qualifications, technical competence and relevant experiences, cultural competence and communication skills are also key attributes of an evaluation team leader.
- The evaluation team should have a clear division of labor among team members by assigning specific roles.
- It is important to stress the importance of stakeholder involvement from the beginning and throughout the evaluation process, thus increasing ownership of the evaluation and utilization of its results.
- Setting a proper evaluation management structure, and sharing and discussing preliminary findings before reports are submitted are important.
- Risks of evaluation findings and recommendations leading to negative effects on beneficiaries or touching on political sensitivities (and thus not getting used) could be mitigated by regular communication with stakeholders during the evaluation process.

**Asia Regional Evaluation Evidence Exchange
October 19, 2015**

SESSION 1: QUALITY IMPROVEMENT IN EVALUATION - CARAVAN SESSIONS

3. Local Capacity Building

Station Manager: Chandan Samal, Project Development Specialist (M&E), USAID/India

Summary: In a practical setting it is rare for managers to be involved in evaluations because they are too occupied with implementation. M&E staff go on the field visits and serve as the liaisons with external evaluation organizations. Some people think this is a missed opportunity to build the capacity of managers, but this separation creates more objectivity. Instead of involving managers directly in evaluations, teams can generate interest in evaluations while building local capacity. This is achieved through evaluation discussion groups inside the organization. Training can also be provided by outside organizations.

Highlights:

- Other organizations in country can be used to conduct evaluations and provide capacity building to an organization's staff.
- Including staff in an evaluation team for a project raises concerns about objectivity.
- Involving regional or headquarters staff in an evaluation can be challenging since that staff member may disrupt the team's work.

Reflection: When thinking about local capacity building, the level needs to be considered 1) institutional level – linking M&E technical staff on management, governance and finance; 2) individual level – understanding the involvement of evaluation staff. It can be useful to create separations between the project implementers and evaluators. This separation enhances objectivity since the project managers maintain some distance from the evaluation teams. However, this limits the ability of an organization to use an evaluation process as a capacity building opportunity for their managers.

**Asia Regional Evaluation Evidence Exchange
October 19, 2015**

SESSION 1: QUALITY IMPROVEMENT IN EVALUATION - CARAVAN SESSIONS

4. Quality Assurance

Station Manager: Yumiko Kanemitsu, UNEDAP Co-Chair, Regional Evaluation Specialist, UN Women

Summary: The process of evaluation is as important as evaluation findings. There are three phases in the evaluation process: Preparation, Conduct and Use. Preparation entails actions such as preparation of evaluation budget and planning, purpose and assessment, stakeholder analysis, evaluation criteria and indicators, as well as assembling an evaluation team and a management structure. When it comes to conduct, we think about the framework, methodology, inception report and methodology, collecting and analyzing data, preliminary findings and reporting. In the Use phase, we focus on dissemination of results and response management.

Highlights:

- The preparation phase in the evaluation design is most critical.
- We must share standard guidance on an evaluation with partners, sharing how we need to collect and analyze the data.
- Stakeholder engagement and consultation is essential.
- A peer-review mechanism can increase quality of evaluation.

Reflection: It's important to have the right evaluation questions and the right team from the very beginning of an evaluation design. To ensure the quality of evaluation, the preparation phase is most critical. There should be standard guidance on evaluation to direct partners on what data they need to collect in the field, as well as a standardized and consistent data collection plan.

Stakeholder engagement and consultation play an important part in the evaluation process. The challenge is defining if it's valuable stakeholder engagement or an interference with the objectivity of the evaluation. In some cases, this type of engagement slows down the evaluation process.

Peer-review is not common in evaluation, but donors are starting to see the need in having experts that can independently provide peer-review of evaluations, especially in specialized methodologies.

**Asia Regional Evaluation Evidence Exchange
October 19, 2015**

SESSION 1: QUALITY IMPROVEMENT IN EVALUATION - CARAVAN SESSIONS

5. Developing a Budget and LOE for an Evaluation

Station Manager: Monica Pons, Deputy Development Program Officer, USIAD/RDMA

Summary: From a donor perspective, when constructing the budget for the scope of work, it's important to figure out who the experts are, what their expenses might be and how they physically get to the sites to do what's been planned. In some case these figures might not be adequate due to lack of local understanding. Issues such as complexity, timing and usability of the data affect the overall evaluation budget.

Highlights:

- Budgets should be determined based on complexity of the project such as partnerships and time allocated for the project.
- Allocate budget contingency for short-term evaluation in case issues come up during implementation.

Reflection: In most case, M&E is a certain fixed percentage of the budget. What needs to be addressed is the complexity of the project, such as partnerships that are difficult to manage. The M&E budget should not be proportional to the overall budget, but to the complexity of the project.

It's important to make a distinction between M&E for programs and short-term evaluation. Budgets can be constrained, but there should be a ten percent contingency fund allocated for issues that come up during the implementation of short-term evaluation. These types of evaluations are precise and if there is a budget adjustment that awaits approval for four to five days, that affects the overall synchronization of the project.

**Asia Regional Evaluation Evidence Exchange
October 19, 2015**

SESSION 2: LEARNING AND UTILIZATION OF EVALUATION

Moderator: William E. Brady, Assistant Dean for Research and Academic Services, Thammasat University

Panelists: Robert Birkenes, Supervisory Development Officer, USAID/RDMA
Stephane Rousseau, Faculty Member, School of Global Studies, Thammasat University
Bruce Ravesloot, Vice President Asia, TANGO International

Summary: Discussion focused on three questions:

1. What is the level of organizational effectiveness in using the findings of evaluations and/or sharing those findings?
2. USAID and the UN have given increased emphasis over the past 5 years on improving evaluation quality and in using evidence to direct our development designs. Has this push improved our project design and/or results?
3. Is it appropriate to employ “classical” M&E approaches in complex environments? Does evaluation accurately reflect this complexity? Are we are keeping up with innovations in technology?

Highlights:

- RDMA has a tracking tool to formalize recommendations, to gauge whether RDMA will accept recommendations, and a timeline.
- USAID did an evaluation of its evaluations. It examined 134 evaluations over 4 years, and interviewed 250 staff. The agency found that 39% of evaluations were used to modify or design projects. When they were not used it was usually because the evaluation was not timely.
- If the purpose of the evaluation is not clear, it will not be used.
- The study of complexity offers a wealth of useful knowledge we can access for M&E. Cascading failure analysis is one example.
- Scaling up is not just about increasing numbers. The difference between pilot projects and scale up requires different indicators and planning.
- Sometimes we add complexity when simplicity is appropriate. Inception phases are an increasing, positive trend, allowing projects to focus M&E resources on specific M&E questions.
- We can't always use classical approaches, since in some places staff can't even get out to talk to people.

Reflection:

- Technology can help us achieve outcomes, but don't just use it for the sake of using it.
- USAID sometimes fails to go back to old evaluations. For example, in some countries USAID has been involved in coffee farming for 40 years and could benefit from revisiting old evaluations.
- M&E can be embedded in a program, but the common monitoring in quarterly reports is not being used to design. Knowledgeable implementers are not being given opportunities to come into USAID and design projects.
- “Pause and reflect” will soon become one of 3 elements of USAID's Collaborating, Learning and Adapting approach that will become a mandatory part of project design.

**Asia Regional Evaluation Evidence Exchange
October 19, 2015**

SESSION 3: LEARNING AND UTILIZATION OF EVALUATION

Facilitator: Jost Wagner, Change Initiative

Presenters: Christopher Frey, Regional Engineering Officer, USAID/RDMA,
Mike Duthie, Senior Director for Impact Evaluation, Social Impact, Inc.

Summary: In the example of the typhoon disaster in Micronesia, looking at technology and available software such as Folcum helped significantly in creating a database for data collection. Due to geographic complexities, being able to access the data from anywhere was crucial. It also enabled sensitivity analysis and quick follow-up responses. The Folcum software played an important part in collecting data for the construction progress report, necessary in the post-disaster context.

In the case of mobile-based remote data collection for M&E, SMS and voice based mobile data collection both share strengths and limitations. The advantages of these types of technologies are:

1. Measuring indicators that fluctuate over time (such as natural indicators, seasonal, etc. that change over time).
2. Increasing the sample size in difficult access requirements, difficult or expensive places to get to or safety concerns.

Some of the strengths of the mobile-based remote data collection are cheaper and faster processes than face-to-face surveys, better and more reliable quality assurance, and safety of data collection. Some of the challenges are: low rate response, sampling and response bias (who owns the phone and who answers the questions), and the limited number of questions asked.

Highlights:

- Technology enhances efficiency when collecting data through mobile phones.
- Measuring the mechanism for change in intervention is difficult, it's something that needs to be measured over time.
- SMS and voice based data collection share strengths and limitations.

Reflection: Mobile phones are used in different contexts, and each context has its challenges and successes. For example, mobile phones were used during an epidemic outbreak to track down the outbreak's specific location. The data for the outbreaks were linked directly to a map. It was an arrangement with the mobile phone providers and free of charge.

Challenges of using SMS survey was addressed in this session. Lack of sufficient credit to send SMS has been the main concerns for low response rate i.e. low response rate received in an SMS survey among 20,000 youths in Uganda. .

Strengths of using mobile phone in the data collection was discussed that it helps to get a database of all the household in a selective area as GPS points for the household. Using mobile phone in data collection also helps to protect confidentiality, which contributes to host answers.

Asia Regional Evaluation Evidence Exchange
October 20, 2015

SESSION 4: HOW TO INCORPORATE CROSS-CUTTING ISSUES IN EVALUATION

Facilitator: Jost Wagner, Change Initiative

Presenters: **Kerry Richter:** Special Professor, Institute for Population and Social Research (IPSR)
Yumiko Kenemitsu: UNEDAP Co-Chair, Regional Evaluation Specialist, UN Women
Elizabeth Davne-Easton: Deputy Program Officer, USAID/Cambodia

Summary: This session discussed the evaluation of health systems. The challenges of producing a complex evaluation was touched upon as it involved government at several levels, CBOs, women's groups, health services, multiple implementers, water/sanitation, advocacy, and a multitude of other elements that were hard to map. The challenges of incorporating gender issues into evaluations was also presented in the session. Variety of frameworks and tools that can be used to measure gender equality, and the methodological biases that exist among development practitioners were discussed. Presentation on the mid-term evaluation of a biodiversity and conservation project was made in this session that it was successful as the technical office and the project implementer took the evaluation seriously and made changes that improved the project.

Highlights:

- Local expertise is essential and it helps to have someone who knows the project.
- The evaluators had to look for patterns that would be universal for the project.
- The India report focused on lessons, what worked, and how things were prioritized.
- The India evaluation would have benefited from 3.5 weeks of field work instead of 2 weeks, which would have allowed the team to reflect more.
- The UN Evaluation Group drafted guidance to harmonize UN evaluation practices. It explains how to integrate a human rights based approach and principles of gender equality, gender analysis, and how an evaluation should be designed. Social indicators (legal, economic, etc.) help to incorporate gender into different types of programs.
- A premise of the rights based approach is that different groups must benefit equally.
- When we evaluate participation and empowerment, quantitative data is not enough.
- Gender specialists can be biased against quantitative methods.
- The evaluation in Cambodia resulted in real changes, like hiring a gender specialist.

Reflection: When working in complexity it can be tempting to simplify, but sometimes you need to understand the complexity before you can simplify. In India the evaluators tried to build a diagram to understand the complexity, but didn't use it in the report. In a project so complex, it would have been good to give the team more time in the field to understand and reflect.

When it comes to measuring gender outcomes there are significant biases from donors and other development actors about everything concerning methodologies and what can be measured. Part of the reason for this is the confusion over when different tools should be used and how they fit into the broader picture of understanding complexity. One of the biases in examining gender issues comes from the fact that gender specialists are usually activists. As a result, they are suspicious of our ability to show causality with numbers. In some cases gender specialists also fail to understand the broader issues a project is addressing. One way to work around this issue is to focus on questions of social inclusion rather than specifically asking questions through a gender lens.

**Asia Regional Evaluation Evidence Exchange
October 20, 2015**

SESSION 5: GOOD PRACTICES IN QUANTITATIVE AND QUALITATIVE METHODS – WORLD CAFÉ

Quantitative Surveys and Peace Polls

Presenter: Pauline Tweedie: Deputy Country Representative, Asia Foundation

Summary: Measuring impact at a quantitative level for programs such as democracy and governance can be difficult. A perception survey is a tool that can help quantify qualitative results. Perceptions surveys were initially designed and used as a programmatic tool and not as an evaluation tool. Establishing the link between the changes is difficult to attribute to programmatic intervention. The level of difficulty depends on the level of intervention. For example, if looking at the overall trajectory about what people feel about a country's major issues, perception surveys can benefit the country strategy.

Highlights:

- A perception survey is a tool that can help quantify qualitative results.
- Perception surveys are accepted as a research method.
- Pre-testing the survey questions at the design phase is crucial.
- Perception surveys help capture levels of knowledge, belief, attitude and opinion.
- Consideration of the reliability of data collection is key.
- It's critically important to have qualitative elements.
- Perception surveys are not meant to measure absolutes.

Reflection: Designing a perception survey is a robust, technical method. Hiring a professional who has done this type of survey before and having a mixed group to help with the questionnaire and to pre-test it is critical. Some of the challenges that come up in perception surveys are translation into another language and sampling. It is important to understand what the smallest unit is, a unit that is able to compare.

In methodology, thinking about the most reliable method and the reliability of data collection is crucial. Face-to-face interviews, telephone-based interviews, and panel surveys are some of the methods used for perception surveys. Using digital data collection ensures a greater level of quality data control; it also helps get better quality data when the response entails sensitive information. A perception survey is never a full-quantitative methodology. It's critically important to have qualitative elements. The types of questions define the survey and those questions are not meant to measure absolutes. They help understand what's causing a certain change.

One of the limitations for perception surveys is collecting data in conflict and fragile situations, where there are physical security issues and either the enumerators or the respondents are in harm's way.

In certain areas such as democracy and governance, violence and corruption, perception surveys are conducted out of necessity, since these are the areas in which it is hard to quantify using other monitoring measurement approaches.

Perception surveys are expensive and they take time. An alternative approach to reducing costs is to have a panel survey, particularly for a long-time program, then inform the sample that you will follow up with a phone call for the follow-up rounds.

**Asia Regional Evaluation Evidence Exchange
October 20, 2015**

SESSION 5: GOOD PRACTICES IN QUANTITATIVE AND QUALITATIVE METHODS – WORLD CAFÉ

2. Qualitative Comparative Analysis

Presenter: Carroll Patterson: Partner, SoCHa, LLC

Summary: Qualitative Comparative Analysis (QCA) is a statistical model that is useful for managing complexity. It forces us to think about changes and conditions, those qualitative changes that answer the question, “so what?” Once the methodology is set up and used, we get the same responses and that means it’s formal, yet at the same time it’s participatory. QCA answers the question, “how?” For example, impact evaluation is about how the project caused the effect. The challenge is that we don’t know what it was about the project that caused the effect. If there are 20 different components in a program, QCA helps tell us which one or which combinations of activities explain the impact.

Concepts:

1. Necessity and sufficiency - necessity means you must be present for the outcome to occur, and sufficiency means there might be multiple pathways to achieving the outcome.
2. Configurational thinking - thinking in terms of combinations of factors, rather than a single one. What external factors are necessary?

Limitations:

- QCA is data driven - if it can’t be modeled, it can’t be done.
- The unit of analysis should be as low as 10 or 20 cases.
- There can’t be any holes in the data.
- Applying an average to the data doesn’t work. It’s not about an average, but a specific pathway.

Highlights:

- QCA helps in managing complexity. It’s important that we identify essential elements of how complex the intervention needs to be.
- QCA identifies when the lack of a certain factor also affects success. For example, a government service delivery program is a success only if it lacks corruption. The lack of corruption is as important to explain how we achieved the element of success.
- QCA requires qualitative anchoring on the data and variation.
- In QCA we can’t infer the opposite.
- QCA looks at interdependency and the intersection of different factors.

Reflection: Statistics can identify combinations, but it does it very differently. Instead of looking at interaction of variables, QCA looks at intersection, which is the combination of how things work in combination with each other. It is more like a division of labor.

In terms of how labor intensive it is to set up the QCA system, it is a participatory process; it doesn’t make judgments on the program. It is an open dialog with the implementers asking them what’s relevant and where their data is, then models their intervention. The timeframe for a QCA depends on how long it takes to set the model, which depends on what data is available or if more is needed for a participatory approach. Then the models are run in a software program. For analysis, it is necessary to talk and investigate further and that is a process that takes time.

**Asia Regional Evaluation Evidence Exchange
October 20, 2015**

SESSION 6: SHARING KNOWLEDGE AND TRATEGIC LEARNING

Moderator: Suzanne Polak: Regional Learning, Monitoring and Evaluation Advisor, USAID/RDMA

Panel: Reena Nadler: Knowledge Management & Organization Learning Specialist
USAID/Washington
Stephen Miller, Director, FHI 360
Carroll Patterson, Partner, SoCha, LLC

Highlights: Summary: This session focused on the concept of strategic learning and what it means for program evaluation. The discussion touched upon collaborating, learning and adapting (CLA) concept at USAID, the different types of learning and how learning is an extension of M&E. It also focused on the importance of constant learning and building that aspect into the organizational management.

- CLA collaboration means internal collaboration within the mission, and external collaboration between mission staff and implementers, as well as host governments and local stakeholders. Learning is to insure that we are learning from what we are doing and changing. Adapting is informing programmatic decisions, learning about the context.
- CLA is a broader concept than knowledge management. It asks what the learning agenda is and if we are completing the knowledge cycle. It includes not only processes for learning, or how we are collecting knowledge, or what we're doing with it, but also if we're enabling the environment that fosters a learning organization.
- CLA is taking processes such as M&E, project design, and strategy design that are outlined in the program cycle and making them more learning oriented, more adaptive and reflective of the environment in which we operate.
- CLA incorporates change management into the concept of becoming a learning organization, as well as prioritization of time, resources and thinking about realistic expectations.
- Learning is evidence based M&E. It is the evidence that is produced by M&E and taken forward into program design.
- Some evaluations are not utilized after they are produced, because the program has gone forward and the timing was not right.
- Development management was created to incorporate development expertise, content, network and social skills for the situation where awareness in a country context is necessary.
- Learning should be built into the organizations to close the gaps such as high staff turnover.
- We need to convince senior management to fund learning within the organization.
- E-communities of practice are a way of learning from peers.

Reflection: The development manager is a concept that some organizations are interested in and trying to implement. For some, it started with the culture within the organization, in particular with the leadership. It opened the door to a new ground of new possibilities of illustrating where they want to go. One of the things that came out of it was the leadership charter, which is not for leaders in particular, but for everyone in the mission who consider themselves leaders. It also focuses on developing expertise and bringing them back. It opens the door to new methods and new learning.

One of the key challenges in an organization is the lack of understanding that it's an investment. The link between the individual learning and institutional learning is frequently separated. Institutional learning has to come from individual learning.