

---

GOLDILOCKS DEEP DIVE

# Organizational Challenges of Impact Evaluation



Copyright 2016 Innovations for Poverty Action. Organizational Challenges of Impact Evaluation is made available under a [Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License](https://creativecommons.org/licenses/by-nc-nd/4.0/). Users are free to copy and redistribute the material in any medium or format.

**FEBRUARY 2016**

**Acknowledgments:** The Goldilocks toolkit was authored by Mary Kay Gugerty, Dean Karlan, and Delia Welsh, with editing and design by the IPA Communications Team (David Batcheck, Laura Burke, Jennifer Cowman, Heidi McAnnally-Linz, Megan McGuire).



**ipa**  
INNOVATIONS FOR  
POVERTY ACTION

# Organizational Challenges of Impact Evaluation

Organizations face many challenges in measuring their impact using rigorous evaluation methodologies such as randomized controlled trials (RCTs). While technical requirements—such as sufficient sample size—are often an issue, for most organizations the main challenge is overcoming internal and external resistance to impact evaluation.

Some objections are well placed. For example, most operational learning does not require impact evaluation. And for some organizations, a focus on measuring impact may not be appropriate, particularly when it takes resources away from other strategic priorities and program monitoring. External pressures from donors or other stakeholders can be the driving force behind impact evaluations, with both positive and negative consequences. When there is little regard for accuracy or quality, organizations waste resources that could have helped to improve their programs.

Below we describe four of the main organizational challenges to impact evaluation beyond technical requirements, and discuss under what conditions these challenges can be overcome or resolved.

## Challenge #1: Aligning Incentives

Aligning incentives around impact evaluation is perhaps the most important challenge to getting an impact evaluation off the ground. Programs and donors face a number of incentives to *not* rigorously evaluate their programs. Most immediately, staff involved in implementing a program may resist impact evaluation because they feel threatened by the premise that the program may not work. Measuring a program's impacts is risky for staff—the organization may learn that the program is not effective and stop implementing it. From a social and resource perspective, it is undoubtedly better to find this out and shift resources to more effective programs, but this may be of little comfort to staff who have worked hard to develop a program. As a result, it is common to find that resistance to impact evaluation often comes from those who are most closely connected to program implementation.

Donor incentives may also present challenges on several fronts. First, donor organizations may have incentives to scale a program before good impact evidence is available, rather than wait for the results of an impact evaluation. Or donors may not want to learn that charitable funds went to an intervention that was not effective. The board of directors of an organization may hold similar views on impact evaluation.

## Conditions that help overcome the challenge:

### Separate Accountability from Learning and Impact

Impact evaluation is a learning tool and often not well suited to demonstrating accountability. If donors and managers are interested in accountability, the organization's M&E system should focus on whether the program was responsibly implemented and tracked. This type of monitoring is likely to include measuring and reporting the inputs, outputs, and, in some cases, the immediate outcomes of the program.

On the one hand, a program may be well implemented (resources were used appropriately in program activities and resulted in designated outputs, such as schools built or teachers trained)

from an accountability perspective. On the other hand, this same program may not demonstrate impact: the outputs may not lead to changes in the well-being of beneficiaries (students are not better educated and more employable). By separating accountability from impact, managers are not assessed on aspects of a program beyond their control.

### **Focus Impact Evaluation on Important Strategic Questions**

Impact evaluation should answer a strategic question for the organization. Answering a question that is deeply important to the organization also ensures that the results will be actionable and affect the design of the program. At the same time, many organizations need help with putting in place a process to identify and prioritize evaluation questions that are the most relevant and actionable for the program.

### **Consider Impact Evaluation as an Opportunity for R&D**

Impact evaluation is well suited to testing pilot interventions or add-ons to existing programs, such as new products or ways of implementing a program. When there is an opportunity for an organization to pick one of two or many courses of action, incentives for conducting the evaluation are more likely to be aligned.

### **Reframe Donor Expectations**

It may be possible for grantees to push back or change donor expectations about measuring impact when they are not consistent with grantee learning objectives. For example, discussions could center around the type of evaluation results that would help the grantee to improve the program and also help the donor make decisions about supporting similar programs.

### **Checking for These Conditions**

To ensure that evidence from an impact evaluation is not set aside or quickly forgotten, an organization should create a plan of action before an evaluation and commit to it before seeing the final results. This may involve discussing several plausible results, and for each result writing down what action an organization intends to take if that result materializes.

### **Examples from the Goldilocks Cases**

[One Acre Fund](#) applies impact evaluation methods to measure the impact of its program and to test new products.

[Digital Green](#) is conducting a randomized evaluation of its program after investing in measures to improve the quality of activity monitoring data from partners, and is looking to confirm the results of a previous non-experimental study.

## **Challenge #2: Making Tradeoffs**

Non-profits, social enterprises and donors all have a limited amount of resources they can devote to measuring their programs. The Goldilocks principle of Responsible emphasizes that the benefits of data collection need to outweigh the costs. In other words, organizations have to make tradeoffs about the resources devoted to M&E versus the intervention itself. Organizations also need to make tradeoffs about what can be measured *well* in any impact evaluation.

In general, data-collection costs are a big ticket item in impact evaluation cost. However, even when data costs are minimized, most impact evaluations, especially randomized controlled trials (RCTs), need careful shepherding by M&E and program staff. For example, program staff may need to

devote time to working with the evaluation team as they jointly develop the research questions and the study design, create a budget,<sup>1</sup> and determine the data requirements. Program staff are often responsible for ensuring that the study design (e.g., the separation of treatment and control groups) is respected, and they may help with data collection. If the evaluation requirements threaten to overwhelm organizational capacity, then an impact evaluation may not be the right fit until capacity issues can be addressed. If internal capacity is less of an issue, the organization will still need to balance the amount of resources that go to monitoring and performance reviews against those spent measuring impact.

## **Conditions that help overcome the challenge:**

### **Senior Management Buy-In**

When getting credible evidence of impact is a high priority for an organization's senior management and key stakeholders interested in scaling the program (such as donors or decision-makers), it will be easier to build internal buy-in for the evaluation. Senior managers may also be able to build these priorities into the organization's culture, helping to ensure that buy-in for evaluations persists even when individual evaluation champions move on.

### **Outside Resources to Support Evaluations**

Funding opportunities for rigorous impact evaluations have increased in recent years. For committed organizations that lack the resources to conduct an impact evaluation, it may be useful to look for external financial assistance. Potential sources include multinational organizations, government, development banks, UN agencies, foundations, and research initiatives within evaluation organizations such as Innovations for Poverty Action, J-PAL and the International Initiative for Impact Evaluation (3ie).

### **Targeted Evaluation Support**

Donors interested in developing evidence in a specific sector or around a specific intervention may support organizations willing to add a research component to their program's implementation. For example, the research initiatives at IPA and JPAL offer competitive funding for research in Financial Inclusion, Entrepreneurship, Agricultural technology adoption, Peace building, and Urban Services.

### **Checking for these Conditions:**

Often the first step in building institutional buy-in is for managers to participate in training or executive education courses on how to apply evidence from impact evaluations. Other indications of senior management buy-in are the existence of strategic plans or other major planning documents that mention impact evaluation and its role in helping the organization achieve the strategic planning goals, or annual budgets that include resources for impact evaluation (indicating priority setting).

---

<sup>1</sup> Common budget categories include: staff (principal investigator or lead researcher, program evaluation staff, including research manager, research assistant, statistician and field coordinator), travel costs (local and international travel), data collection costs (survey design and pilot, field office maintenance and supplies, hiring and training field staff, quality controls, survey material and equipment, transportation to and from the field), data entry and cleaning costs, and analysis and dissemination activities (Glennerster, R., & Takavarasha, K. (2013). *Running Randomized Evaluations*. Princeton Press).

## Examples from the Goldilocks Cases

[Root Capital](#) has an operating model that makes credible impact evaluation difficult to accomplish operationally. Root Capital is working with an external evaluator to explore credible evaluation options, but in the meantime, Goldilocks recommends a cost-benefit analysis for farmers working with Root Capital businesses, rather than impact evaluation with a counterfactual.

[Acumen](#) has set aside funding in its Lean Data Initiative to help selected investments conduct impact evaluations that they could otherwise not afford.

## Challenge #3: Asking the Right Question

A good evaluation question is one that seeks to confirm or better understand a program's theory of change and generates actionable information for decision makers. The challenge of asking the right evaluation question is tightly linked to the challenges of aligning incentives and making tradeoffs. An impact evaluation must meet the technical requirements for a well-designed evaluation, and answer an important strategic question for the organization. Developing the right evaluation question is not always straightforward and may be a contentious process.

Agreeing on the right question and outcomes to measure requires a well-established theory of change that clearly lays out the program elements and intended outcomes. The theory of change is necessary for two key reasons. First, a clear theory of change tells a story about how the program turns resources and activities into outputs and, theoretically, outcomes and impact. Mapping this process is necessary to identify the right question to evaluate, and informs the study design. Second, any new impact evaluation should expand knowledge about the effectiveness of a given intervention. A critical component of the Transportable principle is that the intervention itself must be well understood—as laid out in the theory of change—so it can be compared to existing evidence.

Since any new evaluation should expand knowledge about the intervention, decision makers need to review available evidence before undertaking an evaluation. This step is more than just an academic exercise; it is important for ensuring that resources are used well. Literature reviews can be especially helpful in identifying the most relevant research question, and can also help organizations define the main outcomes to measure. For smaller or new organizations, an evidence review often requires knowledge or capacity that the staff may not have. Fortunately, several organizations have been surveying the [evidence landscape and defining evidence gaps](#).

Once an organization has decided that an evaluation will generate transportable knowledge, staff must agree on the outcomes to be evaluated and work with researchers to develop an appropriate study design. Again, the theory of change is critical to this process. For many organizations, embarking on an impact evaluation may be the first time that a formal theory of change has been developed for the program; the process of developing the theory and agreeing on questions for evaluation can be challenging. Deciding which questions can be answered with the available resources means making hard tradeoffs; rarely will an evaluation answer everyone's "most

important” question. This highlights the importance of aligning the incentives and expectations of staff, boards and donors.

## **Conditions that help overcome the challenge:**

### **Senior Management Buy-In**

Senior management supports the development of a research agenda and the organization has started the process of developing a theory of change and an agenda for knowledge generation.

### **Internal Analytical Capacity or Other Relevant Resources Exist**

An organization has the internal or external capacity to compare the existing program to the evidence available and determine whether to conduct an evaluation. Usually, the decision to conduct an evaluation is made when evidence reviews uncover a lack of evidence or credible evaluations on similar programs. It may also occur when the evidence review indicates impact or program components that require further study.

### **Checking for these conditions**

Use the [systematic review resource](#) to assess whether there is relevant evidence for the program. In some cases, it may be possible to connect with research organizations like IPA and J-PAL, or funding organizations like 3ie, to request information on new evidence or research in progress.

### **Examples from Goldilocks cases**

[Women for Women International](#) is considering impact evaluation of its program, which could contribute to growing evidence around social protection programs that offer a package of services including support for income generating activities.

[Splash](#) has institutional buy-in for rigorously evaluating the impact of its program, but has determined that the organization should carefully consider whether there are relevant WASH evaluations that have been completed or are underway which address the most relevant research questions for the Splash model.

## **Challenge #4: Find the Right Timing**

Choosing the right time to evaluate an idea or intervention can be tricky for many organizations and evaluators. The evaluation results will be most useful in making decisions about expanding or replicating a program, or changing important aspects of how the program is delivered. Evaluations conducted while the program is still being designed or when the proof of concept is not yet certain, however, risk measuring the wrong thing or underestimating what the program impact would be if the program were implemented with full fidelity. On the other hand, evaluating a program that is near or at full scale is difficult because evaluators are unlikely to find a sufficiently large sample of non-beneficiaries to create a valid counterfactual.

An impact evaluation is more likely to produce useful results when beneficiaries participate in the program or use the product as anticipated by the theory of change, and when there is reasonably consistent implementation where the impact evaluation will take place.

Another aspect of the timing challenge is getting results from impact analysis in time to make decisions about program design. Data entry, cleaning, and analysis often take more time than expected. To ensure that analysis takes place in a timely manner, the organization should discuss

their timing needs with the research team, and set expectations on both sides from the start. The logistical requirements of data collection and analysis need to be factored into plans to use the evaluation results for making decisions about continuing or scaling a program.

### **Conditions that help overcome the challenge:**

#### **Project Maturity**

The project is beyond the proof of concept phase, stakeholders can consistently communicate how the program will be implemented, and there is demand for the program (i.e. design and implementation are well-defined and take-up is strong). This does not mean that organizations should not conduct proof of concept evaluations when designing programs; however, these evaluations are likely to be more operational in nature (i.e. addressing questions about take-up and use) rather than measuring the main outcomes of the program.

#### **Consistent Implementation**

The implementation team is able to monitor implementation to ensure consistent delivery.

#### **Timing**

The evaluation question is one that can be answered in a timeframe that matches the timing of decision around continuing or expanding the program.

### **Examples from the Goldilocks Cases**

[Women for Women International](#) has a well-established program that is consistently implemented across program sites. An impact evaluation could help it determine which aspects of its program are the most effective in helping the marginalized women it serves.

## **Conclusion**

Satisfying the technical requirements for measuring impact of a program—the ability to gather credible data and use a credible method—is a necessary but not sufficient condition for investing in impact evaluation. Not all impact evaluations that can be done should be done.

For an impact evaluation to be a responsible investment for the organization, a number of conditions must be met:

- 1) The research question should be of high-priority to the organization, with a strong organizational consensus around the need for impact evidence;
- 2) A review of available evidence should demonstrate that there are knowledge gaps worth filling;
- 3) The organization needs to demonstrate a commitment to acting on the results of the evaluation;
- 4) The program should be sufficiently mature with a strong M&E system in place;
- 5) The costs of conducting an evaluation should be justified relative to its benefits and both costs as well as benefits should be clearly defined.