

---

GOLDILOCKS TOOLKIT

# Goldilocks: Finding the Right Fit in Monitoring & Evaluation



Copyright 2016 Innovations for Poverty Action. Goldilocks: Finding the Right Fit in Monitoring & 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



**ipa**  
INNOVATIONS FOR  
POVERTY ACTION

# Goldilocks: Finding the Right Fit in Monitoring & Evaluation

*The struggle to find the right-fit in monitoring and evaluation (M&E) resembles the predicament Goldilocks faces in the fable “Goldilocks and the Three Bears.” In the fable, a young girl named Goldilocks finds herself lost in the forest and takes refuge in an empty house. Inside, Goldilocks finds a large number of options: comfy chairs, bowls of porridge, and beds. She tries each, but finds that most do not suit her: the porridge is too hot or too cold, the bed too hard or soft – she struggles to find options that are “just right.” Like Goldilocks, organizations have to navigate many choices and challenges to build data collection systems that suit their needs and capabilities. How do you develop data systems that fit “just right”?*

Over the last decade and a half, nonprofits and social enterprises have faced increasing pressure to prove that their programs are making a positive impact on the world. This focus on impact is positive: learning whether we are making a difference enhances our ability to effectively address pressing social problems, and it is critical to wise stewardship of resources.

However, it is not always possible to measure the impact of a program, nor is it always the right choice for every organization or program. Accurately assessing impact requires information about what would have happened had the program not occurred, and it can be costly and difficult (or even impossible) to gather that information.

Yet nonprofits and social enterprises face stiff competition for funding, and to be competitive they often need to prove they are making an impact. Faced with this pressure, it has become common for organizations to attempt to measure impact even when the accuracy of the measurement is in question. The result is a lot of misleading data about what works.

Efforts to measure impact have also diverted resources from a critical and often overlooked component of performance management: monitoring. When done well, monitoring furthers internal learning, demonstrates transparency and accountability to the public, and complements impact evaluations by providing clarity on how program activities are actually taking place.

Simply put, the push for evidence of impact has led many organizations to make three types of mistakes in their data collection efforts:

- **Too little:** Because of the focus on impact, too many organizations have lost sight of what should be their top priority: using data to learn, innovate, and improve program implementation over time. This often requires collecting better data on what an organization is actually doing, and on whether people are actually using their service.
- **Too much:** Getting data has never been easier or cheaper. Cellphones, GPS devices, and Wi-Fi have made it less expensive to gather and transmit data, while a myriad of software innovations has made it easier to analyze and use data. But some organizations collect more data than they actually have the resources to analyze, resulting in wasted time and effort that could have been spent more productively elsewhere.
- **Wrong:** Many organizations track changes in outcomes over time, but not in a way that allows them to know if the organization *caused* the change to happen, or if it just happened

to occur alongside the program. This distinction matters greatly for knowing whether to continue the program.

We put forth four key principles to help organizations of all sizes build strong systems of data collection for both activity monitoring and impact evaluation. The principles are *credible*, *actionable*, *responsible*, and *transportable*, or CART:

- *Credible* – Collect high quality data and analyze them accurately.
- *Actionable* – Commit to act on the data you collect.
- *Responsible* – Ensure the benefits of data collection outweigh the costs.
- *Transportable* – Collect data that generate knowledge for other programs.

## Credible

*Collect high quality data and analyze them accurately*

The principle of credibility involves two parts: data accurately measure what they are supposed to measure and the analysis produces an accurate result.

For data to measure accurately what they are supposed to measure, they must meet three criteria: valid, reliable, and unbiased.

Being “**valid**” means the data (if collected appropriately) should capture the essence of what organizations are seeking to measure. Developing a good measure of a concept can be tricky. Year of birth, for example, gives you a valid measure of age, but often the concept organizations want to measure is broad, such as an attitude towards leaders, or use of health services. We can then ask many questions that fall into the right category, but are imperfectly capturing the essence of what organizations seek to measure. There also should not be alternative interpretations of the data that capture an entirely different concept.

**Reliability** implies that the same data collection procedure will produce the same data repeatedly. There is almost always some randomness to survey questions, and so it is not possible to have perfectly “reliable” data in this sense, but certainly some methods produce less randomness than others, and are thus more reliable.

Measurement of the data should be **unbiased**. Measurement bias refers to the systematic difference between how someone responds to a question and the true answer to that question. Bias can come from many different sources. Respondents can be too embarrassed to report their behavior honestly, leading to systematic over-reporting (or under-reporting) of certain activities. Respondents may not know the answer to a question, or may have incentives to inflate (or hide) some kinds of information, such as income or assets.

The second part of the credibility principle is appropriate analysis. Credible data analysis involves understanding when to measure – and just as importantly – when *not* to measure impact. Even with high quality data, the impact of a program cannot accurately be measured without also accurately measuring what would have happened in the absence of the program.

## Actionable

*Commit to act on the data you collect*

Even the most credible data are useless if they end up sitting on a shelf, never used to help improve programs. Nonprofits today have access to more data than ever before. Electronic data collection, and efficient storage systems allow for much more data collection at a lower cost. In theory, more information should help organizations make better-informed decisions. But in reality, the availability of more data often leads organizations to become overwhelmed by data.

The actionable principle seeks to roll back this problem by calling on organizations to only collect data that they will use. Organizations should ask two questions of each and every piece of data that they want to collect: 'Is there a specific action that we will take based on the findings?' and 'Do we have the resources and the commitment required to take that action?'

The actionable principle can also help organizations decide whether it is worthwhile to conduct an impact evaluation. The rule here is the same as for data collection: organizations should only spend time and money conducting an impact evaluation if they are committed to using the results. Therefore, crafting an actionable evaluation means both designing it in such a way that it will generate evidence that can improve the program but also making an honest commitment to use that evidence, regardless of the results.

## Responsible

*Ensure the benefits of data collection outweigh the costs*

The responsible principle helps organizations weigh the costs and benefits of data collection activities to find the right fit for their organization. Collecting too much data has a real opportunity cost. The money and time organizations spend collecting that data could be used elsewhere in the organization.

On the other hand, too little data can also have societal costs. It is irresponsible to implement a program and not collect data about what took place. A lack of data about program implementation can hide flaws that are weakening a program, lead to the continuation of inefficient programs, and prevent funders from knowing whether funds are being used for their intended purpose.

Like the other CART principles, the responsibility principle can help organizations assess tradeoffs in a number of different areas of M&E, for example:

- **Data collection methods:** Is there a cheaper or more efficient method of data collection that does not compromise quality?
- **Resource use:** Is the total amount of spending on data collection justified, given the information it will provide, when compared to the amount spent on other areas of the organization (e.g. administrative and programmatic costs)?
- **Use of respondents' time:** Does the information to be gained justify taking a beneficiary's time to answer?

As with the other principles, the responsible principle also helps an organization decide whether the timing is right for an impact evaluation. An impact evaluation is a resource-intensive undertaking, making it critical to weigh the costs and benefits, including questions such as:

- How much do we already know about the impact of the program from prior studies, and thus how much more do we expect to learn from a new impact evaluation? Is the added knowledge worth the cost?
- Will future decisions, either by this organization, by other organizations, or by donors, be influenced by the results of this study?

## Transportable

*Collect data the generate knowledge for other programs*

The goal of transportability is to generate lessons from M&E that can help others design or invest in more effective programs. This principle is particularly important for impact evaluations, which generate evidence that can be relevant for the design of new programs or can support the scale-up of programs that work.

To transport findings from one program or setting to another, organizations need an underlying theory to help explain the findings. Such theories need not always be complex, but they should be sufficiently detailed to guide data collection, and to help set boundaries and conditions under which the results are likely to hold and less likely to hold. When organizations are clear about their theory of change and their implementation strategy, it helps others who are doing similar work, either internally or at other organizations.

Replication is a second, complementary method of addressing transportability. There is no better way to find out if something will work somewhere else than to try it somewhere else. Sometimes seeing an intervention work the same way in multiple locations can bolster the policy relevance of the overall set of results.

## Conclusion

For many organizations, adhering to these principles means in some cases they should “just say no” to measuring impact. If nonprofit organizations use existing evidence to inform their programs, when it’s available, and dedicate more energy to measuring what they actually do, they will use their resources more responsibly and generate knowledge that enables both programmatic learning and operational improvement.

Ultimately, Goldilocks seeks to make data collection and decision-making more useful, approachable, and feasible. The hope is that by adding clear guideposts into the messy field of monitoring and evaluation, each organization will find the path that works for them.