Impact Evaluation Notes No. 2. April 2012

LINKING MONITORING AND EVALUATION TO IMPACT EVALUATION

Burt Perrin

This is the second guidance note in a four-part series of notes related to impact evaluation developed by InterAction with financial support from the Rockefeller Foundation. The other notes in this series are: *Introduction to Impact Evaluation*; *Introduction to Mixed Methods in Impact Evaluation*; and *Use of Impact Evaluation Results*. The complete series can be found on InterAction's website at: www.interaction.org/impact-evaluation-notes.

InterAction
A UNITED VOICE FOR GLOBAL CHANGE

THE ROCKEFELLER FOUNDATION

Photo: Manuel Meszarovits

Contents

Introduction

The purpose of nongovernmental organizations (NGOs) and other development actors is to help people improve their lives, communities, societies and environments. NGOs are typically interested in contributing to social change – such as increased justice or economic and social security – while recognizing that these represent long-term outcomes that go beyond the achievement of short-term results.

As Guidance Note 1 indicates, that is what impact and impact evaluation are all about: systematically and empirically identifying the effects *resulting* from an intervention, be they intended or unintended, direct or indirect. Impacts are usually understood to occur later than – and as a result of – intermediate outcomes. There is an increasing realization that good intentions are not enough. Impact evaluation goes beyond considering what agencies are doing to *what happens as a result of these activities*, and the extent to which these interventions are indeed making a difference in the lives of people and the conditions in which they live.

Most NGOs engage in a variety of monitoring and evaluation (M&E) activities. The extent of these activities can vary considerably. In many cases, what goes under the rubric of M&E consists mainly of monitoring and reporting, although organizations are increasingly also engaged in a variety of evaluation activities, examining various aspects of their performance. Yet there is significantly less attention to the evaluation of impact.

This guidance note will illustrate the relationship between routine M&E and impact evaluation – in particular, it will indicate how both monitoring and evaluation activities can support meaningful and valid impact evaluation, and even make it possible. Impact evaluations are typically external, carried out in whole or in part by an independent expert from outside an agency.[1] Nevertheless, M&E has a critical role to play, such as:

- Identifying when and under what circumstances it would be possible and appropriate to undertake an impact evaluation.
- Contributing essential data to conduct an impact evaluation, such as baseline data of various forms and information about the nature of the intervention.
- Contributing necessary information to interpret and apply findings from impact evaluation. This includes information about context, and data like the quality of implementation, needed to understand *why* given changes have or have not come about and what we can do to make our efforts even more effective in the future.

Section 1 of this guidance note discusses some of the characteristics of impact evaluation and how

1 For example, see the USAID *Evaluation Policy*, January 2011. Nevertheless, impact evaluations can and sometimes are also carried out internally, or in combination. Guidance Note 1 has more to say about who should conduct an impact evaluation.

this differs from monitoring and from other forms of evaluation, and more specifically how these different but complementary approaches can be linked to support meaningful impact evaluation.

Section 2 provides guidance and ideas about the various steps involved and approaches that can be used to maximize the contribution of routine M&E to impact evaluation.

One primary audience for this guidance note is *M&E staff* in NGOs, to assist them in identifying steps that they can take to better link M&E activities to impact evaluation. This note may also be of interest to staff in other types of organizations, such as foundations.

In addition, this note will also be relevant for *NGO management*, as well as for *program staff*. Why? First, as noted in Section 3, there is little point in conducting impact evaluation unless the entire organization thinks in terms of impact and continual improvement, questioning how to improve on what is being done. The organization must be willing to act on implications from impact evaluation and other forms of evaluation.

Second, M&E is not something to be left just to evaluation specialists. Program staff can and must play a very active role in collecting data and recording their observations and experiences, which can help ensure that impact evaluations are addressing the right questions and accurately reflect what programs are doing.

1. How can monitoring and other forms of evaluation support impact evaluation?

Monitoring and evaluation can make essential contributions to impact evaluation. Indeed, *meaningful* impact evaluation[2] is not possible without significant support from an organization's regular M&E activities. While the scope of this note is too focused to discuss the nature and diversity of monitoring and evaluation, it is important to recognize some significant differences between "monitoring" and "evaluation," which make different contributions to impact evaluation. Thus, it is helpful to consider some basic characteristics, including differences and opportunities for complementarities, before identifying more specifically how "M&E" can contribute to impact evaluation.

1.1. Main characteristics of monitoring, evaluation, and impact evaluation

The following table summarizes some of the most common and generally agreed-upon characteristics of monitoring, evaluation and impact evaluation. This table should be viewed with care. As some of the wording (e.g., "typical," "can," "generally") and the text below suggests, there is often some variation in how these concepts are viewed and applied.

Monitoring generally involves tracking progress with respect to previously identified plans or objectives, using data easily captured and measured on an ongoing basis. While monitoring most frequently makes use of quantitative data, monitoring qualitative data is also possible and some agencies do this regularly.

Monitoring is carried out for a variety of different purposes, generally having little to do with impact evaluation. Some of the most frequent reasons for monitoring include:

- Internal use by project managers and staff to

2 Impact evaluation that is "meaningful": would appropriately represent the intervention in question and its context; is valid in all respects (e.g., providing for both internal and external validity); and, above all, provides useful information that can help inform future directions.

Monitoring, Evaluation and Impact Evaluation: Some Basic Characteristics

Monitoring	Evaluation	Impact Evaluation
• Periodic, using data gathered routinely or readily obtainable, generally internal, usually focused on activities and outputs, although indicators of outcome/impact are also sometimes used • Assumes appropriateness of program, activities, objectives, indicators • Typically tracks progress against a small number of preestablished targets/ indicators • Usually quantitative • Cannot indicate causality • Difficult to use by itself for assessing impact	• Generally episodic, often external • Goes beyond outputs to assess outcomes • Can question the rationale and relevance of the program, objectives and activities • Can identify both unintended and planned effects • Can address "how" and "why" questions • Can provide guidance for future directions • Can use data from different sources and from a wide variety of methods	• A specific form of evaluation • Sporadic, infrequent • Mostly external • Generally a discrete research study • Specifically focused on attribution (causality) in some way, most often with a counterfactual • Generally focused on long-term changes, such as in the quality of life of intended beneficiaries • Needs to take into account what was actually done (e.g., through M&E) as well as identify impacts

better understand and track how things are proceeding, mainly to identify if the project is on target or not. This includes tracking data on what services actually are being provided, the quality of services being provided, who is being served, and related considerations. Monitoring data can also often serve as an early warning system, and in the case of negative or unexpected findings may suggest the need to consider a change in approach while the project or program is still underway.

- Internal organizational use at the regional, national and/or international HQ level so the agency can track a project's or activity's status against plans and expectations; for planning and management purposes; and to address accountability needs of the agency's board, funders and the public.
- Addressing external requirements for compliance and control, such as donor demands for reporting and accountability.

Monitoring can also take other forms. Rather than focusing specifically on what the organization is doing, for example, monitoring could include citizens' report cards with respect to public services,[3] advocacy services of many organizations that may track corruption practices, environmental approaches, employment of child workers in local business practices, etc.

---

3 Further reading: Samuel Paul. *Holding the State to Account: Citizen Monitoring in Action.* Bangalore: Public Affairs Centre, 2002.

Some definitions of monitoring:

A continuing function that uses systematic collection of data on specified indicators to provide management and the main stakeholders of an ongoing development intervention with indications of the extent of progress and achievement of objectives and progress in the use of allocated funds. (OECD/Development Assistance Committee [DAC] Glossary)

The tracking of project outputs and outcomes as indicators of project effectiveness, or the extent to which the project achieves its stated objectives. (USAID)

While these are all legitimate and important reasons for monitoring, none are concerned with contributing to impact evaluation. As a result, often the types of data that are collected, and the ways in which these are kept and reported, are difficult to use for evaluation. As the box to the right suggests, one needs to *plan in advance* in order for monitoring data to be useful for impact evaluation. In ideal circumstances, those conducting an impact evaluation can contribute to the design and structuring of an intervention's monitoring system.

At the same time, it is important to acknowledge the limitations of monitoring. Monitoring mainly tracks progress against predefined objectives and indicators, and assumes these are appropriate. But monitoring alone is insufficient for drawing conclusions about attribution, or for identifying the *reasons* why changes have or have not taken place (such as the extent to which these changes are a result of the intervention or due to other causes). It is also usually unable to identify unintended effects, gaps in service, etc. For this, one usually needs evaluation.

An all too common scenario: an impact evaluation planned at the end of a project finds that no baseline information or other appropriate data has been collected or appropriately disaggregated to demonstrate changes or even provide necessary information about program activities, who has been served, and other needed information. This leaves no way to identify impact or attribute it in any way to the program.

Evaluation involves systematic, evidence-based inquiry that can describe and assess *any* aspect of a policy, program or project. Evaluation uses a wide variety of both quantitative and qualitative methods, providing more comprehensive information about what is taking place, why, and whether it is appropriate or not, and to provide guidance for future directions.

Evaluation can be carried out for many different purposes and take a variety of forms. Some of the following types of evaluation (the list is not inclusive – there are also many other evaluation approaches) may contribute to impact evaluation under certain circumstances.[4]

- **Needs assessments** involve assessing or evaluating the needs or problem situation, often prior to the initial development of the project design. Such assessments frequently identify ways in which these needs can be addressed.
- **Process (or implementation) evaluations** describe the nature of the intervention as it is

4 The definitive resource for definitions of key terms used in development evaluation, such as those used in this note, is: OECD/Development Assistance Committee. *Glossary of Key Terms in Evaluation and Results Based Management,* http://www.oecd.org/dataoecd/29/21/2754804.pdf. Also see: Michael Scriven. *Evaluation Thesaurus.* 4th Ed. Sage Publications, 1991. USAID's Evaluation Policy groups all forms of evaluation other than impact evaluation under the category of "performance evaluation."

actually implemented in practice. To a certain extent, monitoring may be able to provide data about program activities that can be useful for impact evaluation. However, interventions are rarely applied exactly as initially intended and frequently change over time, often for good reasons. It can be surprisingly difficult to determine what is actually taking place, who is being served, in what ways or to what extent, and what else is going on. Process evaluation can go into more depth than is possible with monitoring, often explicitly using questions arising from monitoring as a starting point.

Programs frequently need to change, sometimes considerably, in response to information from needs assessments and other forms of feedback from monitoring and evaluation. For example, a program targeting orphans and vulnerable children initially assumed that children living alone would be most vulnerable, but quickly learned that there were households headed by adults where children were even more vulnerable.

Without understanding what "the program" *really* is, even the most sophisticated and statistically rigorous impact evaluation will have little meaning. Evaluations that clearly outline the program, along with reasons for divergence from original expectations, can provide invaluable information to help understand how the program's interventions might have made an impact (or, perhaps, to indicate some challenges in a program's implementation or some of its underlying assumptions that may impede its ability to make an impact). For example, if a program's impacts were limited, data from process evaluation can help ascertain if this was because of a problem in the *theory* of how the program was expected to work, or due to limitations in how it was implemented.

Some definitions of evaluation:

The systematic and objective assessment of a planned, on-going or completed project, program or policy, its design, implementation and results. (OECD/DAC Glossary)

The systematic collection and analysis of information about the characteristics and outcomes of programs and projects as a basis for judgments, to improve effectiveness, and/or inform decisions about current and future programming. (USAID)

- **Formative evaluation,** carried out partway through implementation, is intended to improve performance during the subsequent steps of a program or project. This can help identify intermediate outcomes, at what point (if any) the intervention seems likely to make an impact and what else may be needed to enhance effectiveness. Consequently, this can make for more meaningful impact or summative evaluation[5] carried out at a later date.
- **Organizational evaluation** looks at an organization's overall effectiveness, or perhaps that of an organizational unit. Organizational factors (e.g., governance, management, human resources, finances, intra- and inter-organizational relationships, etc.) often may have more to do with an intervention's success than its design does – these factors are essential information that must be taken into account during the design and interpretation of impact evaluation.

5 Summative evaluation, often equated with impact evaluation, involves assessing the extent to which anticipated outcomes were produced at the end of a planned intervention and making a judgement as to its worth.

In addition, the OECD Development Assistance Committee (DAC), one of the most influential bodies concerning development evaluation, has identified five basic evaluation criteria (or questions): relevance, effectiveness, efficiency, impact and sustainability.[6] Note that evaluation of *impact* represents just one of these criteria. It is rarely possible to conduct an impact evaluation focused specifically on attribution without also having undertaken other forms of evaluation to better understand what has taken place. There is little point in documenting that an intervention is making an impact if it is no longer relevant or if there may be more effective or less costly ways of addressing the basic need, or if the intervention and/or its results are not likely to continue or to be sustainable.[7] Guidance Note 1 notes circumstances under which impact evaluation could be useful.

Impact evaluation generally shares the basic characteristics of other forms of evaluation, and should be one of an NGO's evaluation approaches. However, as the table on page 3 suggests, there are typically some significant differences, underscoring the need to use other forms of M&E to make an impact evaluation meaningful.

Guidance Note 1 observes that impacts and impact evaluation are sometimes defined in different ways. Nevertheless, an essential aspect of impact evaluation concerns *attribution*, the linking of documentable impacts in a cause-and-effect manner to an intervention. Simply knowing that impacts have come about because of an intervention, however, is insufficient. To be able to apply the findings from impact evaluation in other settings and/or to other groups of beneficiaries, one needs to know *why* and *how* the given results came about, as well as the characteristics of those who did (or did not) benefit.

This is where meaningful M&E can be helpful – provided that one identifies in advance, as specifically as possible, the types of information that will be needed and how various forms of M&E can provide these.

1.2. How M&E can contribute to impact evaluation

As the above discussion indicates, there are many types of data that M&E can, indeed *must*, contribute to enable impact evaluation. There are four basic categories of information that M&E can contribute to impact evaluation.

1. Information about the nature of the intervention, which is essential for impact evaluation to causally link any documented outcomes to what was done. Basic information from routine M&E is required about: the actual nature of services provided; baseline data, as well as changes over time, with respect to who is served (or addressed, such as with advocacy initiatives); the number and characteristics of beneficiaries reached, including relevant subgroups; program outputs; and how the intervention has progressed. These may be different from what was originally expected. As previously noted, the planned intervention can vary significantly from what was actually implemented, either in response to feedback from M&E or from changes in circumstances. The targeted population may also change during the life of the program, as organizations learn more about who should be eligible for a program or

---

6 One of many sources with more information about these criteria is the *DAC Criteria for Evaluating Development Assistance*. This and many other DAC evaluation resources can be found at http://www.oecd.org/document/35/0,3746,en_21571361_34047972_31779555_1_1_1_1,00.html.

7 Frequently, interventions subject to impact evaluations or other intensive research are resourced at a higher level than "normal" work, in a way that is unsustainable or that can readily be "scaled up" or applied in other contexts.

become better at reaching them. Without taking this into account, any findings from an impact evaluation may be meaningless.

Much of this data is quantitative in nature and can come from monitoring, along with qualitative descriptions about the nature of the program or project. Process evaluation in particular often can provide more in-depth information about the extent to which services (or messages, in the case of an advocacy program) are actually reaching the intended audiences, how these are viewed and how they vary across different groups of people.

2. Information about the context of the intervention. What other interventions and factors may influence whether or not an intervention can "work"? Can different groups of people be affected differentially (e.g., aid provided after a disaster disproportionately going to those best able to access it; services and other forms of assistance requiring people to go to a site, therefore leaving people with mobility limitations unserved)? These kinds of considerations must be taken into account in the design of impact evaluations.

Similarly, what influencing factors may be special about the context for this intervention? For example, the successful delivery of food aid and subsequent lives saved frequently depends on things that are well beyond the program's control, such as logistics, weather, support (or interference) from governments or terrorist organizations, etc. This kind of information is essential for meaningful interpretation of impact evaluation data, and for identifying the findings' implications for other settings or situations. These types of data are mainly (but not necessarily) qualitative. They sometimes can come from monitoring, but in many cases would require evaluation to enable more in-depth exploration. Guidance Note 3 provides more information about potential methods that can be used to collect these and other forms of data.

3. What we already know (or at least suspect) about impact. Sometimes there already is some evidence – or at least a strong suggestion – from monitoring and/or evaluation that some kind of changes may be taking place. In such cases, an impact evaluation could document and confirm what is really going on and the extent to which these changes have been a result of program activities. In contrast, M&E may raise questions about the viability of an impact evaluation, at least at the present time, indicating that it might still be premature to look for impact. Once an impact evaluation is underway, data from monitoring and other forms of evaluation should of course be taken into account, along with any other data collected.

4. What other evidence may influence the decision to undertake an impact evaluation. As indicated above, evaluation information about an intervention's relevance, efficiency and sustainability should be taken into account before setting priorities for impact evaluations.

2. How to build impact evaluation into M&E thinking and practices

As identified above, M&E is vital to impact evaluation – but its contributions are not automatic. This section provides guidance and ideas for how to plan and carry out M&E so that it *can* support meaningful impact evaluation. These are organized around the steps identified in the diagram below, and discussed in the text that follows.

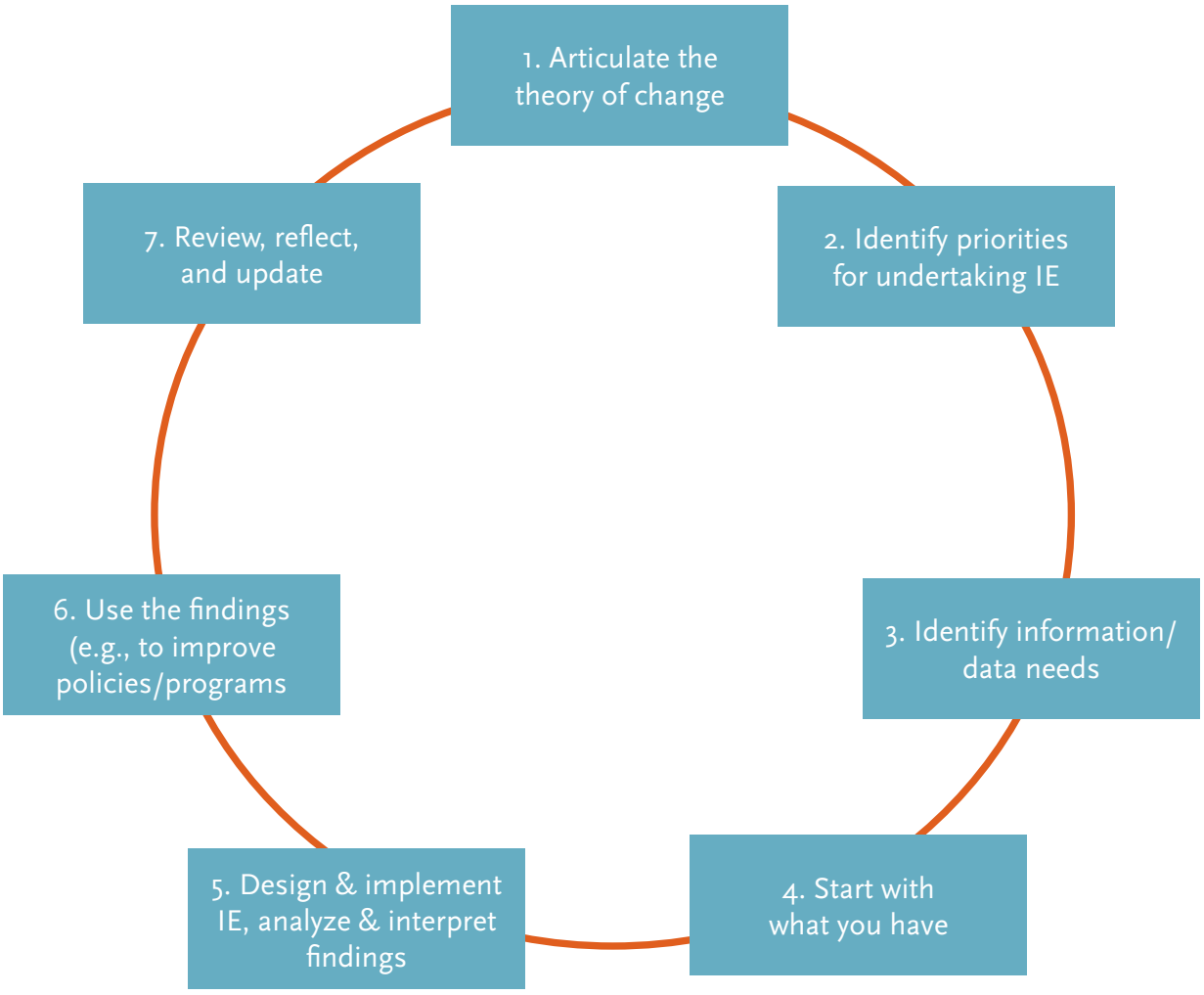
2.1. Articulate the theory of change

As Guidance Note 1 indicates, a *theory of change* is a model that explains how an intervention is expected to lead to intended or observed impacts.

Often referred to as the *program theory*, *results chain*, *program logic model* or *attribution logic*, the theory of change illustrates the series of assumptions and links identifying the presumed relationships between:

- Inputs (e.g., funding, staff, volunteers, tangible and in-kind support from others, etc.);
- Activities and their immediate outputs (i.e., what is done, such as training sessions conducted);
- Intermediate outcomes at various levels (e.g., teachers trained, teachers applying what they have learned in the classroom situation); and
- The intended impact (such as reduction in child labor, families and communities that have become financially self sufficient, improved health status, fewer people dying in emergency situations, etc.).

Guidance Note 1 has identified key elements and various ways in which a theory of change can improve impact evaluation. This, in turn, has some very direct implications for M&E that can be summarized under the following main categories.



A theory of change for advocacy efforts, similar to those for direct service projects, should identify the sequence of intermediate steps (e.g., awareness, other partners agree to take complementary action, community mobilized, dialogue with government officials created) and other intervening factors (e.g., activities by others) by which this would be expected to lead to changes in government policies and practices, or perhaps in community practices.

1. Identify the stage in the trajectory of an intervention when it would be appropriate to undertake various forms of evaluation – including impact evaluation – by identifying the various intermediate steps and other factors that may also influence impact (such as interventions of other allies, contextual variables, etc.). M&E may be able to track progress regarding activities, outputs and initial outcomes, indicating when it could be possible to consider impact evaluation. This can avoid the all too common situation of evaluations that attempt to measure impact prematurely, sometimes even before a program has a chance to get well established.

Michael Woodcock of the World Bank has emphasized the importance of determining the timeframes and trajectories of impact that is reasonable to expect. It is critical to recognize that different projects may follow very different trajectories. As he observes, "Some projects may, of their nature, yield high initial impacts ... while others may inherently take far longer, even decades, to show results, not because they 'don't work' after three years, but because it's simply how long it takes" (Woolcock 2011). It is also important to bear in mind that an organization's staff may be overly optimistic about the time it takes to set up a new program, or for results to follow. To attempt impact evaluation prematurely is to set up a program for failure.

2. Describe the manner in which the intervention is expected to work using a theory of change. Often referred to as the *results* (or *impact*) *pathway*, this can help suggest the types of M&E data needed to document the nature of the intervention and how it has evolved in practice. The theory of change can indicate which aspects of implementation need to be checked for quality, to help distinguish between *implementation* failure and *theory* failure. It also provides a basis for identifying where along the impact pathway (or causal chain) an intervention may stop working. This type of information is essential to draw a causal link between any documented outcomes or impacts and the intervention. It is also essential to explain and interpret the meaning and implications of impact evaluation findings.

3. Develop a common understanding of the intervention by involving staff and other key stakeholders in the development of the theory of change. If a theory of change is developed just by an M&E specialist relying extensively on documentation, there may be no ownership of the model. It also may not reflect the reality of what is taking place. It is advisable, where possible, to involve program, field staff and some other key stakeholders in developing the theory of change. There is a crucial role for the M&E specialist in the process, as an active facilitator and drawing up the actual model. Sometimes there may be differing views about how a program may be expected to "work," resulting in alternative theories of change. This can suggest questions to be explored by M&E, perhaps before undertaking the actual impact evaluation.

In addition, if a participatory approach is taken, the development of the theory of change can help all

participants think in outcome terms. The process can help develop ownership and a common understanding of the program's purpose and what is needed for it to be effective.

2.2. Identify priorities for undertaking impact evaluation

Many factors help establish priorities for undertaking impact evaluations, such as interests of key stakeholders or donors, the availability of funding or opportunities to collaborate with others. But a key consideration is when it would be most likely to contribute useful information that cannot be obtained through simpler and less costly means, and that can provide useful guidance for future strategies or program approaches.

M&E plays an important role in this process. The theory of change can indicate which questions are appropriate to address when during a program. Monitoring and evaluation data can help to indicate how far interventions have progressed, at what stage they are "ready" for impact evaluation, and when there is sufficient supporting M&E documentation to make this possible. Start by identifying information needs, then see how much information from routine M&E is already available before deciding to undertake an impact evaluation. At this stage, one can assemble the performance story to clarify how much is already known, pinpoint gaps in information and unanswered questions, and identify when an impact evaluation would be most applicable. This may be to provide "hard" evidence to confirm what is suspected, or in situations when there are real questions about an intervention's value that cannot be answered through other means.

2.3. Identify information/data needs

What types of M&E data will be needed to support impact evaluation? The simple answer is: "it depends." There is no one-size-fits-all prescription; that would be a surefire route to irrelevancy. This is why this guidance note has emphasized the importance of articulating the theory of change, preferably involving program staff and other stakeholders in the process, and using techniques such as a realist evaluation mindset and a contribution analysis approach as described below. These techniques will help identify questions for evaluation, which in turn can help identify the necessary types of data, and M&E techniques and methods to obtain these.

Guidance Note 1 lists examples of key evaluation questions for impact evaluation. Section 1 of this note identified various ways M&E data can contribute to impact evaluation. But there is no magic formula. It is necessary to identify, very specifically, what types of data will actually be required for impact evaluation, and the extent to which existing monitoring procedures and evaluation practices may need to be modified to provide this information (or to indicate that this is not possible). The following table can serve as a summary checklist of the types of information requirements that M&E approaches should be able to provide in most situations.

Baseline data
• Identify the initial state along dimensions relevant to the intended impact of the program (the theory of change should be able to assist in identifying these dimensions). • In situations where baselines have not been established, are there ways these could be recreated retrospectively?*
Nature of the program/intervention, as *actually* implemented
• Identify changes from how the intervention was initially expected to be delivered, and to the extent possible, the reasons for these modifications. • Identify various modifications and changes along the way (and if there were none, was the initial concept really perfect in all respects, or was the program unable to respond to changing circumstances?). • Identify other descriptive information about the implementation process, including phasing of various activities or variations across sites or types of beneficiaries.
Who did the program serve?
• Disaggregate the beneficiaries' characteristics to the extent possible, and certainly along characteristics identified during development of the theory of change and the performance story. • There are often some differences to track (e.g., gender, age, different tribal/cultural or socio-economic groups) that may be evident, but often others that may be very relevant to a particular situation (e.g., children with or without living parents, living with relatives or on their own).
What else was going on that could affect the program's impact, positively or otherwise, e.g.:
• How has the program worked in conjunction with other agencies or programs, including identified partners? • Have there been other factors – including interventions of other actors, government policies, private sector initiatives (such as the creation of employment opportunities, or perhaps the closure of the major employer in an area), natural and man-made disasters, etc. – that have affected the ability of the program to progress positively or negatively, or perhaps may have required a somewhat different approach than initially envisioned? • To what extent have the above factors helped or hindered the program?
What outcomes or impacts can be documented, in the short-, medium- and long term?
• These should relate to the theory of change, taking into account what types of outcomes were expected at given points in time. • Are these as intended, or not? Any data, or even subjective assessments, as to why this might be? • Are impacts likely to be sustainable? What is the evidence for this? • How strong is the data? What else is needed to make a more convincing case of impact following in some way from the program intervention?
What else happened (unintended/unexpected effects)?
• To what extent were these positive or negative? • To what extent were these potentially under the control of the program? • Should the program design and the theory of change be revised to take these into account for the future?
How appropriate and relevant was the program as designed and implemented?
• E.g., to what extent did the program address beneficiaries' requirements, community or national needs, or other identified local/global priorities? • Have needs and contextual considerations changed that may suggest that the program should be modified in some way?
What else is needed?
• Are there other needs that the program itself should be addressing? • Are there other needs that it might be more appropriate for others to address (e.g., implications for advocacy)?
Other considerations ...
• ... as identified initially in the theory of change. • ... that might arise later, such as in response to preliminary findings, through reflection meetings, etc. • ... that address concerns or perspectives of other stakeholders, including critics of the program.
*** Further reading:** RealWorld Evaluation techniques such as discussed in Bamberger et al. 2012.

Suggestions in this table may well appear as a formidable list of what needs to be monitored. But not every program needs to collect every form of data. The questions need not be asked all at once, or can be staged throughout the lifecycle of the program.

Also, it is important to prioritize information requirements, and to focus on "need to know" information. While developing a monitoring plan, it should be clear in advance how every piece of information can be used. Trying to obtain too much information can create unintended consequences. For example, it may overwhelm the staff who need to collect the data, who are then forced to take shortcuts that can affect the quality and usability of any information provided.

Realist (or realistic) evaluation. When identifying data necessary to support meaningful impact evaluation, it is helpful to bear in mind the tenet of *realist evaluation*,[8] which starts from the premise that saying a program "works" or not is an oversimplification, and that a statistically significant difference between two groups[9] may mask considerable variation. Invariably, some people gain and others do not, and almost all programs may "work" for some people in some situations, but not for others. Realist evaluation states that outcomes arise through various combinations of context and mechanisms. For example, gunpowder will only fire if it is lit in a certain way. Thus, a major implication is that impact evaluation should identify what works for whom, under what conditions and through what mechanisms.

There are many illustrations of approaches (e.g., community mobilization) that need to be adapted based on the livelihoods of the target group, for example pastoralists vs. sedentary populations (such as farmers). Similarly, an approach to improve youth reproductive health may work with some cultural groups but not with others, or be dependent upon other factors.

Realist evaluation principles have major implications for the design of M&E systems, particularly for types of data that need to be collected and kept disaggregated. For example, some program approaches that are effective with boys in certain situations may be detrimental to girls. Some programs that work in urban areas may not work in rural areas. The relevance of some dimensions (for example, gender, age or degree of poverty) may be evident (but still not always tracked), while others may be less obvious. For example, an educational approach that may assist some children may not do so for others, such as children who may have been traumatized in some way. In addition to characteristics of communities and beneficiaries, it is also important to keep track of context, or in this case what services were provided and in what circumstances.

NGO staff, given their close connection to the field and to beneficiaries, are often very well placed to be able to observe, or even suspect, which factors can influence approaches that work for some but not for others. Documenting this type of information through ongoing M&E can be an important contribution to impact evaluation. It is essential, not only to understanding what has actually taken place with a given program, but also to learn from this experience and to be able to apply this information in different situations.

8 Further reading: Pawson and Tilley, *Realistic Evaluation* (1997).

9 Which, statistically, is based on comparing the average scores of each group.

2.4. Start with what you have

Before undertaking an expensive and complex impact evaluation, start with the information already available, from M&E and other sources. It may be that there is already sufficient evidence and an impact evaluation would not be worth the investment. At a minimum, this can help focus an impact evaluation on those questions and issues that cannot be addressed in any other way. It may suggest other forms of evaluation that might be more appropriate to try first. Here are a couple of suggested systematic approaches.

Contribution analysis involves the following six steps, explained in more detail in Appendix 1:

1. Develop the theory of change.
2. Assess the existing evidence on results.
3. Assess alternative explanations.
4. Assemble the performance story.
5. Seek additional evidence.
6. Revise and strengthen the performance story.

In essence, this involves using evidence from existing M&E to see what it can reveal about the outcomes or even impact of an intervention, while also considering what else besides the intervention could have created the identified outcomes. From this, a provisional performance story can be developed about the extent to which it is reasonable to assume that the program's actions could have contributed to the observed outcomes, and to identify weaknesses and where addition data would be useful. This is a powerful way of using existing data to determine what is known and where data is needed from additional forms of M&E, or if necessary from an impact evaluation, to provide a more convincing picture.

Reflective discussions. A contribution analysis approach need not be a major undertaking, particularly for a small and contained project. Systematic reflective discussions (or "review" or "monitoring" meetings) are one good example of this approach. These could form part of regular staff meetings, or perhaps be the focus of special meetings set up periodically. They could involve an informal discussion of what seems to have been working or not; the basis for drawing these conclusions, taking into account competing explanations for any observed or demonstrated outcomes (in other words, a simple version of the performance story); and then identifying what other forms of evidence will be needed in the future to provide a more convincing case of the program's impact. In some cases, such as when staff and stakeholders from different areas have been brought together, it might be appropriate to use a more formalized approach and an external facilitator. This kind of qualitative monitoring can be invaluable in understanding what may have led to outcomes or impacts (and whether these same results would have occurred anyway in the absence of the project), which is essential information in order to apply the findings of impact evaluations.

AWARD, a fellowship program for women scientists in agricultural research and development in Africa, has been using periodic reflection sessions to give the fellows and their mentors in different countries an opportunity to consider progress, achievements to date, and what has and has not been going well. These add to the management team's monitoring data, as well as their own ongoing reflections on performance. This has provided opportunities to test the program's theory of change, and has also helped the management team to work in a "continuous improvement" mode, identifying the need for any remedial steps or changes in direction.

Eliminate rival plausible explanations. Step 3 of the contribution analysis approach involves assessing alternative possible explanations. This is a simple but powerful concept, well anchored in the evaluation literature. It was first articulated by Donald T. Campbell,[10] who further said that one should use the simplest and least costly approaches that can provide necessary confidence in the findings.

Some approaches to ruling out rival plausible explanations

- Anticipate potential questions of the audience(s) for the evaluation.
- Address threats to internal and external validity.
- Use a theory-based approach.
- Identify what types of evidence would be convincing.
- Provide multiple lines of evidence
- Ensure face validity.*

*This refers to findings that are presented in such a way that they appear to make sense at face value to non-researchers

First identify other plausible explanations for how impact may have come about other than the program or project intervention, and then produce the necessary evidence to rule out the likelihood of these alternative explanations. It only needs to address other *plausible* explanations. One can safely assume that changes in youth employment patterns in a district were not caused by the ghost of Elvis! But the opening or closing of a new business operation, or changes in the policies of an existing operation, might be relevant in some situations. It is important to realize, however, that stakeholders can differ on what they view as plausible or not, and that this is something that should be taken into account when developing an M&E approach.

2.5. Design and implement the impact evaluation, analyze and interpret the findings

As previously indicated, M&E has an important role to play in making sure that an impact evaluation asks the right questions and that the design is realistic about what relevant M&E data might exist or could be practically provided. Complementary evaluation exercises involving both qualitative and quantitative approaches may be appropriate, contributing to a mixed methods approach (discussed in Guidance Note 3) that could provide for triangulation of impact evaluation findings and later to assist in interpretation. M&E specialists undoubtedly would need to be involved, at least at some stages during the impact evaluation process, to facilitate additional data collection. M&E specialists also need to be fully aware of issues and tradeoffs involved in establishing any form of validity.[11]

A common assumption is that once an impact evaluation design is drawn up, everything must remain static, often for a period of years. The reality, however, is that programs do evolve and change over time, and indeed must do so in order to be responsive to changed circumstances, feedback and new information. Impact evaluation designs need to be able to reflect and adapt to changes in the nature of the intervention, otherwise the findings may have limited meaning. As Guidance Note 3 indicates, some approaches to impact evaluation can be more flexible and adaptable than others. Nevertheless, it is important to bear in mind that impact evaluation requires cooperation from the program in order to succeed, and to the extent possible significant changes should be discussed with the evaluation team.

10 For example: Campbell, D. T. and J. C. Stanley. 1963. *Experimental and Quasi-Experimental Designs for Research*. Boston: Houghton-Mifflin. Campbell is perhaps best known as the father of modern program evaluation and a proponent of experimental and quasi-experimental methods. He was the first to identify the latter concept, as well as with the concepts of internal and external validity.

11 A very useful very recent resource is: Huey T. Chen, Stewart I. Donaldson, and Melvin M. Mark (eds.). *Advancing Validity in Outcome Evaluation: Theory and Practice*. *New Directions for Evaluation*. Jossey-Bass and American Evaluation Association. 2011.

M&E specialists (and through them, other organization staff and other stakeholders) should play a significant role in the interpretation of impact evaluation findings. It is useful to ask if the findings make sense and are consistent with other information. If not, it would be appropriate to ask what this might mean, and what could explain any apparent contradictions.

As Section 1 has emphasized, M&E is needed to provide information about relevant contextual factors and the nature of the intervention, which will be essential for drawing any causal links. In particular, M&E is needed to provide *explanations* for findings, to indicate the "whys" and "hows," essential information to act upon impact evaluation findings. Are assumptions made about "the intervention" still valid? Counterfactual designs are often particularly unable to explain the differences between experimental and control groups. Even if there is no difference between the mean (average) scores of the experimental and control groups, it is likely that some people nevertheless have done better, and it is still possible that a majority of people could do better. Complementary M&E information can help get inside the "black box" in order to be able to understand and interpret the differences – and what they mean for future directions.[12] Such considerations might suggest additional subgroup analyses that may still be possible, or at a minimum aid in appropriate interpretation of the findings.

2.6. Use the findings

Use of impact evaluation is the focus of Guidance Note 4 and is largely beyond the scope of this note. Nevertheless, it is worth noting that M&E can have a significant role to play in this process. Other M&E information may be able to help identify the extent to which an impact evaluation's findings can be generalized or may be applicable to other settings or populations, or even at the same place, given the inevitable changes in context, the nature of the intervention, and the target group over the lifespan of a multiyear impact evaluation. M&E specialists can help animate sessions involving staff and management to consider implications of impact evaluation findings in light of other information.

2.7. Review, reflect, and update

No matter how much effort one has put into the initial articulation of the theory of change, developing an evaluation plan and design(s), and putting data collection mechanisms in place, this does not mean that one can or should then forget about it. It is useful to periodically consider whether the assumptions underlying the impact evaluation approach are still valid, or if any changes may be needed. Responsive programs rarely remain static – they adapt their approaches to changes in context, to address emerging needs and opportunities, and in response to feedback from beneficiaries, other stakeholders and M&E. If identified early enough, it may still be possible for an impact evaluation to take these into consideration. Appropriate impact evaluation methodologies should be chosen that are flexible enough to accommodate changes in program direction that, in many cases, are almost inevitable.

It is appropriate to conduct interim reviews (which can be as simple as a focused staff meeting) to consider any relevant changes to the program, and to revisit the theory of change to check if the original assumptions are still valid or need updating. As suggested above, this review can be an

12 For example, impact evaluations are often undertaken with the objective of indicating if a given approach should be continued and/or replicated in a different setting (generalizability or external validity). But without knowing how and why impact occurred (or not), it is impossible to make informed decisions about such choices.

opportunity to develop or to update the performance story, and to identify gaps in information that should be addressed in some way. Interim reviews, in addition to providing an opportunity to reflect on how the program seems to be progressing, can also serve as an opportunity to review, and if necessary update, the M&E plan itself.

This note has emphasized the role that M&E can play to support meaningful impact evaluation. But the reverse is also true. Impact evaluation can invariably suggest types of data that should be monitored in the future, as well as raise questions that can be addressed by other forms of evaluation. It is appropriate, at the conclusion of an impact evaluation, to consider implications for future M&E priorities.

3. Engaging all parts of the organization

M&E and impact evaluation clearly require an active role for M&E specialists to put in place appropriate evaluation designs, and to ensure that all aspects of the process are undertaken in ways that maximize the validity and utility of the data. But there are also important roles for management and program staff that sometimes are not well appreciated.

3.1. M&E: A core management function requiring senior management leadership and support

M&E, and thinking and acting in impact terms, are much too important to be left just to M&E specialists. First and foremost, M&E represents a management tool that can help in setting directions; assessing progress; learning about the types of approaches that appear to work or not in varying circumstances; making decisions; and in many other ways. This means that NGO senior management must take ownership, provide leadership, and encourage everyone within the organization to think in terms of outcomes or results by asking how they can know their work is making a difference and what can be done to improve. Learning and change imply an openness to acknowledging and accepting that, on occasion, some approaches will have not worked well and require modification. But this requires management to encourage and support staff to innovate and not to punish staff if not everything turns out as expected – provided that there is openness to learn from experience, and from evaluation. Such support is essential if impact evaluation, along with M&E, will have any real impact on how programs are implemented.

Managers at all levels in an agency will need to provide the necessary support, resources and leadership to enable meaningful M&E – and impact evaluation – to take place. Senior managers in particular have a key role to ensure that this is not just a paper exercise that can breed cynicism and passive compliance, but to demonstrate their commitment to doing – and to using – all forms of results information. Carrying out impact evaluation, even if there is a dedicated budget for this, will also have workload and possible resource implications for program staff and M&E specialists. Management should be aware of these implications when agreeing to undertake impact evaluation. To a large extent, the relevance and success of impact evaluation, and the ability of routine M&E to contribute to this success, is dependent upon the attitude and support of senior management.

3.2. An active role for program staff is required

Meaningful impact evaluation, together with routine M&E, requires the active involvement of program staff. In most cases, it is program staff who will be responsible for the actual collection, recording, and reporting of much of the data. It is important to acknowledge this key role and what it implies. Imposing requirements on busy program

managers and staff is in practice often looked on as yet another administrative task of dubious value and meets with resistance – which may result in issues with respect to data quality and completeness.

Making M&E and impact evaluation helpful to program staff rather than viewed as a chore

- Involve them in the process.
- Address questions that they feel are relevant and can help them.
- Provide opportunity to participate in the interpretation of M&E/impact evaluation data.
- Provide feedback when data is submitted (even an acknowledgement or a "thank you").
- Provide recognition for staff who assist with impact evaluation.
- Share information about how the data has been used.

As suggested above, appropriate leadership from an organization's senior management is needed to set the stage for meaningful data collection. To facilitate the needed cooperation of program staff in data collection, they should be given incentives and the resources needed (time, staff and funding) to ensure the collection and recording of quality data, along with appropriate quality control mechanisms. To make the data collection relevant and meaningful, program staff should be engaged in the process of identifying questions and issues that they feel are important and relevant to them and to the communities in which they are working. The accompanying box lists some ideas for making M&E, and impact evaluation, meaningful. For example, staff in an education-for-all project in Uganda began to appreciate the importance of the monitoring data they were asked to provide, and to improve its quality, once some of the initial evaluation reports were shared with them.

To the extent possible, considerations regarding the role and support of program staff should be built into the M&E/impact evaluation approach. Program staff who will be required to collect the data are often best placed to identify which information is easy to obtain and which may be more problematic, and what requests for data are ambiguous and can be interpreted in differing ways, or may even be so unreliable as to be useless. Through involvement in the process, M&E is more likely to be viewed as a support rather than a burdensome chore with no value, or something that is "done" to them with no discernable value.

Summary

Impact evaluation is sometimes seen – by some researchers as well as by many program managers and staff – as a very technical exercise that can only be carried out by external experts working at some distance from the program, its staff and its constituency. And there are some situations where this is how it is approached.

But as this note has indicated, impact evaluation represents just one form of evaluation. Only in rare circumstances can meaningful impact evaluation be carried out without drawing on data from other ongoing M&E activities. But the value and use of M&E to impact evaluation is not automatic. Advance planning is needed to identify the types of data that will be required in order to be able to determine the impact of policies, programs and projects and how these data can be obtained. And as this note has emphasized, when everyone within an agency thinks in impact or results terms, this alone can help support the effectiveness of the agency's operations – and help improve the lives of people around the world.

References and Other Useful Resources

Bamberger, M., Rugh, J. and Mabry, L. (2012). *RealWorld Evaluation: Working Under Budget, Time, Data, and Political Constraints* (2nd ed.). Sage Publications.

Chen, H., Donaldson, S., Mark, M. (eds.) (2011). *Advancing Validity in Outcome Evaluation: Theory and Practice. New Directions for Evaluation.* No. 130. Jossey-Bass and the American Evaluation Association.

Funnell, S. and Rogers, P (2011). *Purposeful Program Theory: Effective Use of Theories of Change and Logic Models*. Jossey-Bass Publications.

Jones, H. (2011). "A guide to monitoring and evaluating policy influence." London: Overseas Development Institute. http://www.odi.org.uk/resources/docs/6453.pdf

Mayne, J. (2001). "Addressing attribution through contribution analysis: Using performance measures sensibly", *Canadian Journal of Program Evaluation*, 16(1), 1-24.

Mayne, J. (2011). "Contribution Analysis: Advancing Cause and Effect." In *Evaluating the Complex: Attribution, Contribution, and Beyond,* edited by Robert Schwartz, Mita Mara, and Kim Forss. London: Transaction Publishing.

OECD/DAC. (2011). *Evaluating Impacts: An Overview*. http://www.oecd.org/document/17/0,3746,en_21571361_34047972_46179985_1_1_1_1,00.html

OECD/DAC. (2002). *Glossary of Key Terms in Evaluation and Results Based Management*. http://www.oecd.org/dataoecd/29/21/2754804.pdf

Malhotra, A. (2010). "Connecting Routine M&E with Impact Evaluation," International Center for Research on Women. Presentation at InterAction/3ie Impact Evaluation Design Clinic, Washington, D.C., April 2010. http://www.interaction.org/sites/default/files/18/Anju_Malhotra_-_Connecting_Routine_M&E_with_Impact_Evaluation.pdf

Paul, S. (2002). *Holding the State to Account: Citizen Monitoring in Action*. Bangalore: Public Affairs Centre.

Pawson, R. and Tilley, N. (1997). *Realistic Evaluation*. Sage Publications.

Scriven, M. (1991). *Evaluation Thesaurus* (4th ed). Sage Publications.

Woolcock, M. (2011). "Guest Post on The Importance of Time and Trajectories in Understanding Project Effectiveness." Development Impact. Blogs.worldbank.org. http://blogs.worldbank.org/impactevaluations/guest-post-michael-woolcock-on-the-importance-of-time-and-trajectories-in-understanding-project-effe.

Annex 1 – Contribution analysis

John Mayne, then with Canada's Office of the Auditor General, has suggested using contribution analysis as one approach by which one can make best use of existing data about outcomes and impact. He has summarized this approach as follows. More detailed information about a contribution analysis approach can be found in Mayne (2001), from where this table has been copied verbatim, or in Mayne (2011).

Step 1: Develop the results chain	Develop the program theory model/program logic/results chain describing how the program is supposed to work. Identify as well the main external factors at play that might account for the outcomes observed. This program theory should lead to a plausible association between the activities of the program and the outcomes sought. Some links in the results chain will be fairly well understood or accepted. Others will be less well understood or subject to explanations other than that the program was the "cause." In this way you acknowledge that attribution is indeed a problem.
Step 2: Assess the existing evidence on results	The results chain should provide a good idea of which intended results (outputs, intermediate and end outcomes) could be measured. What evidence (information from performance measures and evaluations) is currently available about the occurrence of these various results? The links in the results chain also need to be assessed. Which are strong (good evidence available, strong logic, or wide acceptance) and which are weak (little evidence available, weak logic, or little agreement among stakeholders)?
Step 3: Assess the alternative explanations	Outcomes by definition are influenced not only by the action of the program but also by external factors — other programs, as well as social and economic factors. In addition to assessing the existing evidence on results, there is a need to explicitly consider the extent of influence these external factors might have. Evidence or logical argument might suggest that some have only a small influence and that others may have a more significant influence on the intended results.
Step 4: Assemble the performance story	With this information, you will be able to set out your performance story of why it is reasonable to assume that the actions of the program have contributed (in some fashion, which you may want to try and characterize) to the observed outcomes. How credible is the story? Do reasonable people agree with the story? Does the pattern of results observed validate the results chain? Where are the main weaknesses in the story? There always will be weaknesses. These point to where additional data or information would be useful. If getting additional evidence is not possible (at least for now), then this is the most you can say about the extent to which the program has made a difference.
Step 5: Seek out additional evidence	To improve your performance story you will need additional evidence. This could involve information on both the extent of occurrence of specific results in the results chain and the strength of certain links in the chain. A number of strengthening techniques that you might be able to adopt are outlined in this work.
Step 6: Revise and strengthen the performance story	With the new evidence, you should be able to build a more credible story, one that a reasonable person will be more likely to agree with. It will probably not be foolproof, but will be stronger and more credible.