USE OF IMPACT EVALUATION RESULTS

David Bonbright, Keystone Accountability

Contents

Introduction

The state of current practice with respect to using impact evaluation is at best emergent. As we found in a recent survey of InterAction member evaluation practices, however, impact evaluation use is not significantly different from how organizations use other evaluative activities such as performance evaluations, formative evaluations, developmental evaluations, cost effectiveness studies, and case studies.

Since impact evaluations focus on results and what contributes to results, logically there should be a greater enthusiasm for making use of the findings. Impact evaluations also tend to cost more than other types of evaluation, implying that they should command significant follow up as well.

It turns out, however, that credible evidence and significant costs are not sufficient to ensure that impact evaluation findings are used.[1] This guidance note looks to several additional factors that contribute to effective evaluation use.

Using impact evaluations depends importantly on how one designs and conducts them. As was illustrated in Guidance Note 3 on mixed methods, the choice of evaluation method should follow from the questions one is trying to answer. But whatever questions are being tackled, whatever evaluation methods are employed, to make the best and proper use of evaluation findings we can apply a common set of practices and insights.

Research for this note included an online survey of current impact evaluation utilization practices among InterAction members and follow-up telephone interviews with survey respondents and other evaluation specialists who volunteered to speak in-depth about their work. A report of the survey data is available on InterAction's website as a companion publication to this note.[2] A list of interviewees is provided at the end of this note in the References section, which also includes a number of academic articles and other publications relevant to evaluation utilization. It should be noted that these published references are neither illustrative nor comprehensive of what is a very large and diverse body of work. While the purpose of this guidance note is not an academic one, a review of the literature suggests that it is consistent with the current research. That said, research does not confirm all the recommended practices in this note – at least not yet.

It is a healthy signal that no one engaged through the research for this note is particularly happy with the current state of the art. There is a general sense that impact evaluation can be useful and should be

---

1 Some of the considerable academic literature confirming this point is included in the reference section at the end of the guidance note.

2 The survey was sent to members of InterAction's Evaluation and Program Effectiveness Working Group and had a response rate of 16.1 percent. The survey report can be found at: http://www.interaction.org/document/survey-report-how-we-are-using-impact-evaluations.

undertaken more often. There is a strong appetite to improve the delivery of evaluative activities in general and impact evaluation in particular. Survey and interview participants provided many examples of increasing human and financial investments in evaluation generally. As an area of emergent practice, it is a time of experimentation to discover what will prove to be most effective.

1. Impact Evaluation Use: Three Themes

This note is organized around three themes.

Guidance Note 1 offers a definition for the utility of impact evaluation that echoes the first two of these themes. Themes one and two are closely interrelated and speak to the practical steps that enable use. The third theme considers the wider set of institutional incentives and rewards for using impact evaluation.

Theme one states that use does not happen by accident. It should not depend upon serendipity or luck. As Samuel Goldwyn, the Hollywood movie producer noted, "The harder I work, the luckier I get." Evaluation use depends mainly on preparation. Impact evaluations are more likely to be used when uses have been anticipated and planned from the earliest stages of the evaluation and, even better, from the planning stages of the work that is being evaluated.

This idea runs through all the guidance notes in this series. When impact evaluations are designed at the early planning phases of a development program or intervention, they can improve the theory of change, set the learning questions, enrich the stakeholder engagement strategy, and inform ongoing monitoring design. Planning for the uses of impact evaluation compels us to consider the different audiences for evaluation findings.

Theme two is a natural extension of the first theme and concerns the operations and systems required in an organization to use impact evaluations well. Evaluation challenges most frequently identified by NGOs point to the need for what might be termed a joined up approach across most parts of the organization. This ties in closely to proven practices that support organizational learning. This theme also considers the causal connection between how an evaluation is funded and how it is used. And it also looks at relationships. Setting expectations by engaging and communicating early and often with stakeholders and audiences for the evaluation is critical, as is timing. A significant evaluation finding that is released too late or too early may be a finding that misses its impact. An evaluation's impact turns out to depend a great deal on how well and how quickly its findings can be expressed in terms of strategic priorities and communicated in appropriate ways to specific stakeholders at the time when they need it.

> **Utility** – good impact evaluation is useful. The likely utility of an evaluation can be enhanced by planning how it will be used from the beginning, including linking it to organizational decision-making processes and timing, being clear about why it is being done and who will use it, engaging key stakeholders in the process, and then choosing designs and methods to achieve this purpose.
>
> **— Guidance Note 1** (p. 14)

Theme three builds from the premise that the first two themes are necessary but insufficient conditions for the effective and widespread use of impact evaluations. Unfortunately, one of the biggest barriers to using evaluation findings is a misalignment between the desired behaviors that we could all readily agree to (such as using valid evaluation findings to improve performance) and

the incentive systems that support (or discourage) their adoption. The findings from impact evaluations – and indeed any evaluative activities – will not be used well unless and until we tackle and reform organizational culture. The most challenging suggestions in this note speak to the task of eliminating disincentives and creating incentives for adopting evaluation findings. Without rethinking the incentive and reward systems that guide the behavior of employees, evaluation is likely to remain a marginal activity as opposed to a core driver of decision-making. Theme three sets out directions and principles for this incentive system rethink, with some guiding illustrations from current practice.

1.1. Where are we now?

The online survey on impact evaluation utilization asked respondents to rate their degree of agreement with 10 different statements about impact evaluation. **Figure 1** on page 4 arrays the 10 sets of answers by clustering the respondents into three groups: those who could be considered to be in strong agreement (*promoters* – scores of nine and 10); those who are fairly neutral on the question (*passives* – scores of seven and eight); and those who tend to disagree (*detractors* – scores of zero-six). The net promoter score (NP score) indicated for each is simply the percentage of promoters minus the percentage of detractors.[3]

One might argue that designating those who provide scores of five and six as detractors is possibly overstating disaffection. There is a strong body of evidence in the customer satisfaction industry, however, that suggests otherwise. In this case, moreover, we are taking into consideration

---

3 NP score is fast becoming the standard tool in business for managing customer relationships in part because its summary of the views across a population is more useful than means and medians.

the low response rate to the survey – 16.1 percent. In the craft of measuring customer satisfaction, many include non-respondents (in this case 83.9 percent) as detractors. We have not done that. We do believe, however, that it is reasonable to assume that: (i) those who responded to the survey have a greater interest and investment in impact evaluation than those who did not respond to the survey; and (ii) those who responded tend to view impact evaluation more favorably than those who did not. Thus, it is reasonable to conclude that the attitudes arrayed in **Figure 1** are more positive towards impact evaluation than the attitudes among NGOs as a whole.

The only statement that has strong support is that impact evaluation does no harm. The largest number of detractors manifested for "My organization emphasizes the use of impact evaluation findings." The reasons most commonly pointed to for this in the open responses included a lack of understanding of impact evaluation concepts, the absence of the necessary operations and systems to use findings, and funding constraints.

We also asked respondents to tell us how the organization would characterize itself with respect to impact evaluation if it had to choose one of five descriptions: opponents, doubters, qualified supporters, supporters, and true believers. The results are consistent with the Figure 1 findings. As expected, given that those who responded to the survey are among the more positive towards impact evaluation, no one declared himself or herself an opponent. Half of the organizations characterized themselves as qualified supporters, 42 percent as supporters and eight percent as true believers. Given the competing demands for time and the complexity involved in undertaking and using impact evaluations, combined with the absence of strong incentives or rewards for doing

Figure 1. Attitudes towards impact evaluation organized into promoters, passives and detractors
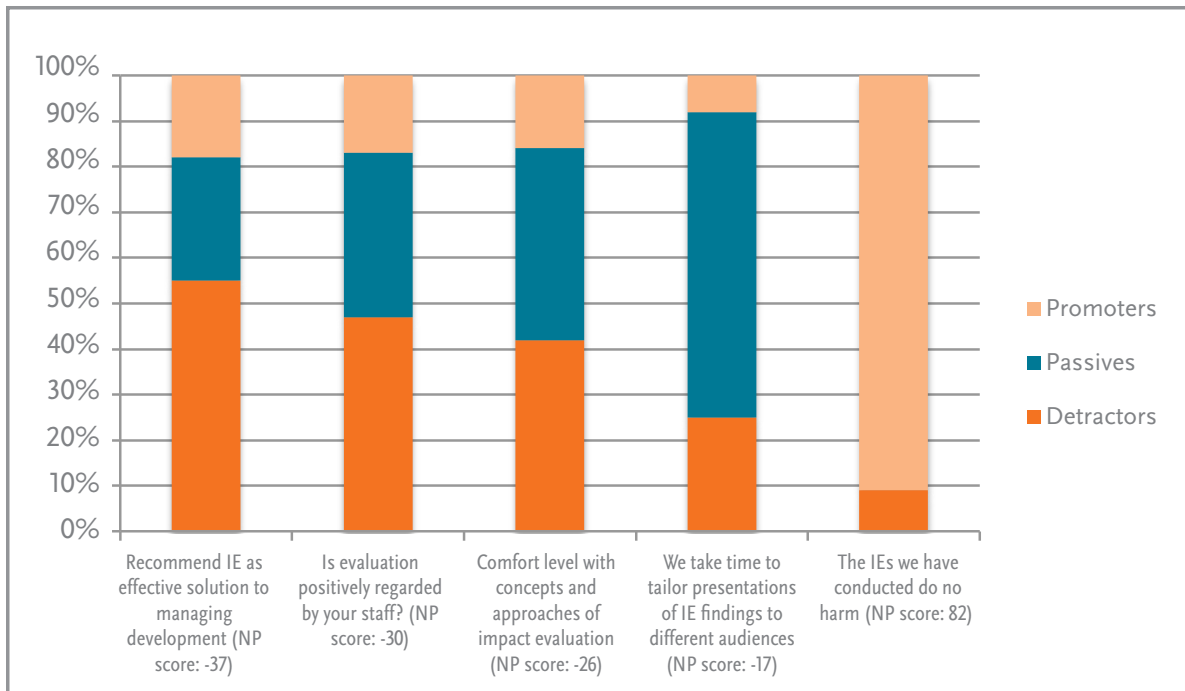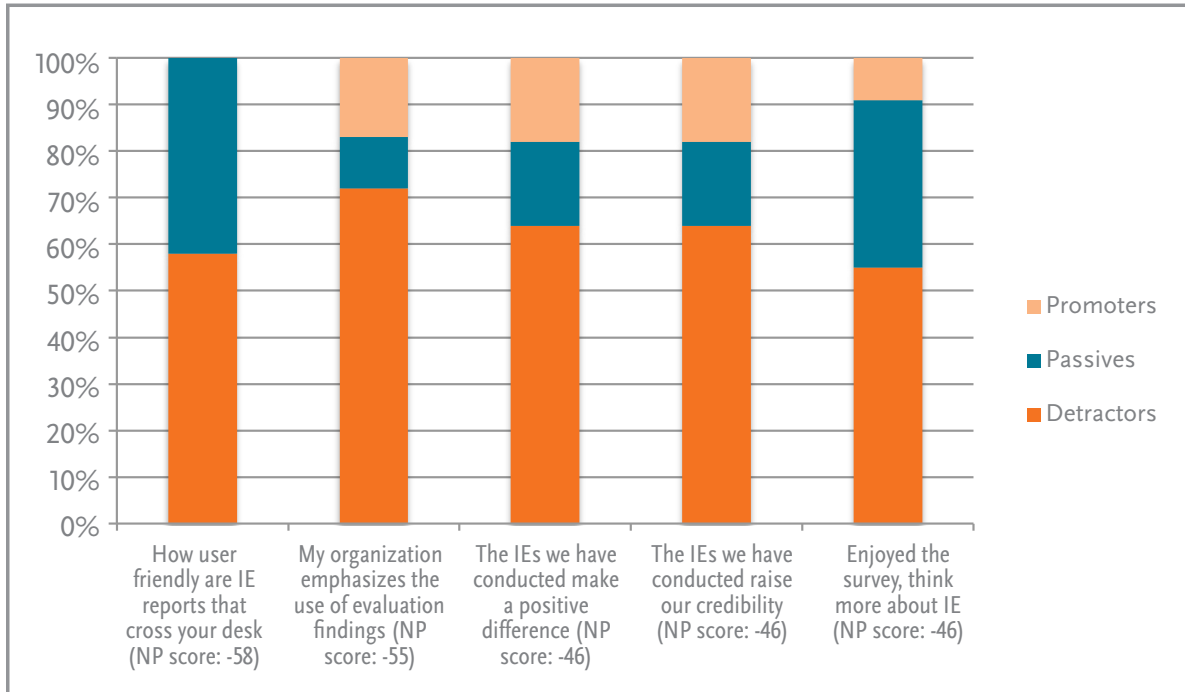
so, it is reasonable to assume that only the true believers are likely to be using impact evaluations effectively.

For organizations with a culture of effective evaluation use, one would expect to see at least half of NGO staff describing themselves as true believers and fewer than 10 percent as self-declared qualified supporters. Organizations that set themselves to reach such targets are the ones who will make the most progress in using evaluations well.

1.2. Recognizing the constraints

Before turning to a detailed discussion of our three solution themes, it may be useful to summarize the commonly mentioned constraints to effective evaluation use.

Five types of constraints are frequently mentioned.

First, many are not clear about the value and uses of impact evaluation. The value proposition is not clear, while the financial costs are significant. Guidance Notes 1–3 respond to this obstacle.

Second, most organizations do not know how to do impact evaluations, let alone use their findings. This is a technical skill that is new to the sector and only coming into common practice now. Guidance Notes 1–3 respond to this obstacle.

Third, organizations do not have the systems or practices to use impact evaluations. At best, they are looking for help to build the capacity to become impact evaluation champions. This guidance note tackles this challenge, especially in theme two, below.

The final two types of constraints are closely related and touch on the wider environment.

In especially candid moments, organizations speak of the potential risks associated with the possibility that evaluations will generate "bad findings," that is, findings that show that the organization is not having an impact, or worse. The main worry here is that negative evaluation findings will lead to reduced funding.

Closely related, donors typically are, if anything, at a more primitive stage in their own understanding and use of evaluation than their grantees. Surveys consistently show that donors neither provide sufficient funding for nor understand how to support and use evaluation well.[4] It is important to clarify that this characterization is of donors as a whole, encompassing individuals and institutions. There are a few high performing exceptions.

When addressing these constraints NGOs (and foundations) most often speak of the imperative to create a culture of learning in which failure is redefined as a learning opportunity, various learning-friendly practices and commitments are recommended, and the features of an organizational culture based on learning and continuous improvement are defined.[5] Another dominant theme emphasized in the professional literature is the importance of leadership. Leadership support is, of course, a *sine qua non* for organizational change and therefore is important to realize effective evaluation use. Conversely, as effective use of evaluation

---

4 One 2007 survey that corroborates this point can be found here: http://www.keystoneaccountability.org/node/493. See also, David Bonbright and Betsy Schmidt, "Taking evaluation seriously: still a ways to go," *Alliance Magazine*, Volume 12, Number 4, December 2007. These 2007 findings were corroborated in a 2012 study by the Center for Effective Philanthropy: http://www.keystoneaccount-ability.org/node/494. Please also see New Philanthropy Capital's "Making an impact: Impact measurement among charities and social enterprises in the UK," http://www.thinknpc.org/publications/making-an-impact/.

5 See, for example, "Four Essentials for Evaluation," Grantmakers for Effective Organizations (2012). http://www.geofunders.org/storage/documents/2012_geo_evaluation_essentials.pdf.

becomes more common, our understanding of good organizational leadership will necessarily come to include effective support for evaluation and its use. In the future, one will not be described as "a good leader" if one is not an evaluation enabler and a systematic user of evaluation findings.

2.   Theme One: Use is for Users

First comes purpose, or more commonly purposes, since most impact evaluations are striking a balance of multiple purposes. Every purpose has built-in implications for particular users.

Given the cost and effort involved in conducting impact evaluations, as well as the comprehensive array of evidence they gather, it often makes sense to consider how different aspects of the evaluation evidence will be useful for particular users. It is highly likely that different users will find greater value in particular parts of the overall evaluation. For example,

the study of a clinical program may wish to ensure that evidence of patient experiences are carefully communicated to clinicians, while findings about attributed health outcomes may be of greatest interest to policy makers, funders and legislators.

The table below compiles two usefully nuanced ways to think about purpose from Guidance Note 1. These approaches to purpose begin to suggest the context in which an analysis of user interests and needs can be undertaken.

Given clarity of purpose, the next step is to list all of the affected stakeholders. These are potential evaluation users, both internal and external to the organization. Most programs would want to consider staff, board, intended beneficiaries, implementation partners (including where relevant other NGOs and government), funders, and policy makers. Many would also put peers, academia/ universities, media and society-at-large on the list.

Why do impact evaluations?	When are impact evaluations most useful?
To decide whether or not to continue or expand an intervention.	Interventions where there is not a good understanding of their impacts, and better evidence is needed to inform decisions about whether to continue funding them or to redirect funding to other interventions.
To learn how to replicate or scale up a pilot.	Innovative interventions and pilot programs that, if proven successful, can be scaled up or replicated.
To learn how to successfully adapt a successful intervention to suit another context.	Periodic evaluations of the impact of a portfolio of interventions in a sector or a region to guide policy, future intervention design and funding decisions.
To reassure funders, including donors and taxpayers (upward accountability), that money is being wisely invested.	Interventions with a higher risk profile, such as a large investment (currently or in the future), high potential for significant negative impacts or sensitive policy issues.
To inform intended beneficiaries and communities (downward accountability) about whether or not, and in what ways, a program is benefiting the community.	Interventions where there is a need for stakeholders to better understand each others' contributions and perspectives.

The next task is to indicate which evidence and findings from the impact evaluation are likely to be most valuable to each stakeholder. These can be stated in the form of propositions or questions that should subsequently be shared and discussed with representatives of that stakeholder group. As Michael Quinn Patton has noted, one needs to clearly and concisely articulate how different stakeholders are likely to use the evaluation information.[6] An evaluation designed to inform those intended to benefit from a program about those benefits, for example, would consult with the intended beneficiaries about the nature and description of purported benefits. Those consultations would seek the beneficiaries' responses to an initial listing of benefit types. Importantly, they would also seek advice about the best way to organize and present the results from the evaluation.

Given finite resources, it is necessary to set priorities for stakeholder engagement, and to discover the highest value opportunities. Including a cost-benefit analysis as part of the planning process will help to win organizational support for engagement activities. You need to answer the questions: For this evaluation, with these purposes, which stakeholders are the most important users? What do they have at stake in the evaluation? How powerful are they in relation to other stakeholder groups? You will need to make provisions for support for low-power groups who have a high stake in the evaluation. Consider rotating more intensive consultation efforts across stakeholder groups over time and through different evaluations so that all groups eventually have their turn to be engaged more intensively. Record the intensity of engagement with different stakeholders over time – by name and group whenever possible – and manage it to ensure all groups are meaningfully included.

6 See, for example, chapter 3 in *Utilization-Focused Evaluation* (4th Edition). SAGE Publications, 2008.

For many if not most evaluations, the primary audience is staff. One independent evaluator tells the story of a potentially contentious evaluation of an international program within a larger, primarily domestically operating NGO. The organization had been implementing the program for a long time under the same leadership, and the vice president in charge of the program was quite committed to it. Knowing that it might be difficult for the organization to accept results indicating that changes to the program were needed, the evaluator began by asking the international program team and CEO how the evaluation results should be organized and presented to be most useful. The evaluator also made it a point to consult the vice president and his team about the scope of the evaluation, and took the extra precaution of intermittently sharing emerging data with the team. This was done both to make sure that the evaluator's interpretation of data was valid, and – importantly – to enhance the team's ownership of the findings. In the end, the evaluation did include some challenging findings implying the need for significant changes. The evaluator believes the team was subsequently able to act on the recommendations because it had been engaged throughout the evaluation.

Initial consultations with users should also discuss: (a) what forms of evaluation report would be most useful for that user; and (b) any time sensitivity related to the evaluation's possible results.

The process of engaging with evaluation users about purported program benefits or other aspects of an evaluation implicitly sets expectations about the evaluation. It is always better to be as explicit as possible about what the evaluation is and is not likely to do, and to articulate a clear process by which the evaluation will communicate its findings.

Whenever possible, once the investigation stage is complete and analysis has begun, evaluation findings should be shared in a preliminary mode, for deliberation and interpretation with the relevant stakeholders. Evaluation evidence normally presents

itself as data that can be interrogated for wider significance. Why are behaviors changing in one case, but differently in another? Good evaluations often generate more questions than they answer. These questions arise from the data and can provide for a highly energizing conversation with stakeholders. If discussed in a preliminary mode, stakeholder interpretations can be included in the final report, making the evaluation much more relevant for users, and providing social validation of the findings.

> "We undertook this rigorous research after consulting with you and taking into consideration your priorities and concerns. This is the evidence that we collected. This is what we think it means. These are some important questions that it raises for us. What do you think?"

The pattern we are describing here is a circular movement from consultation to investigation to consultation and back again.
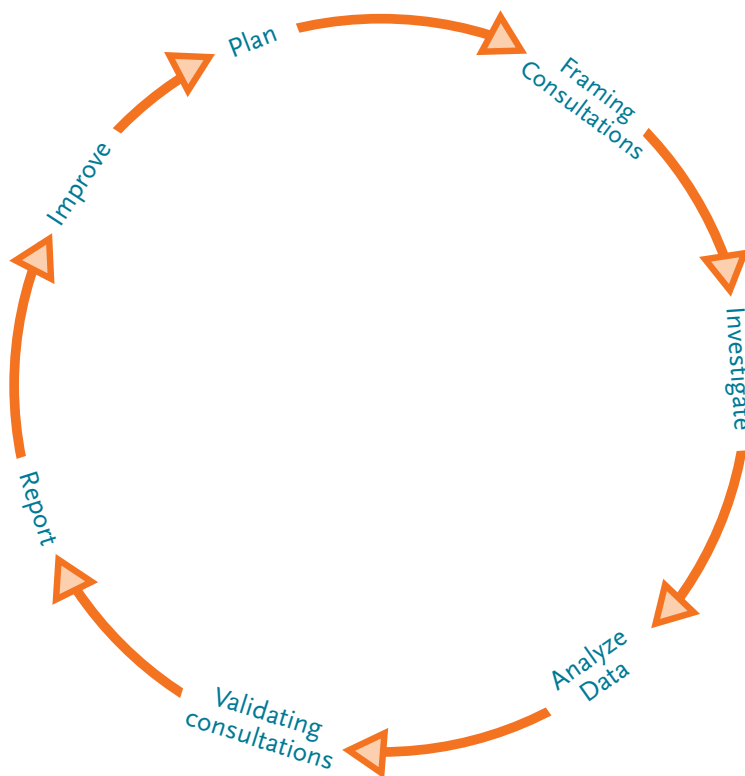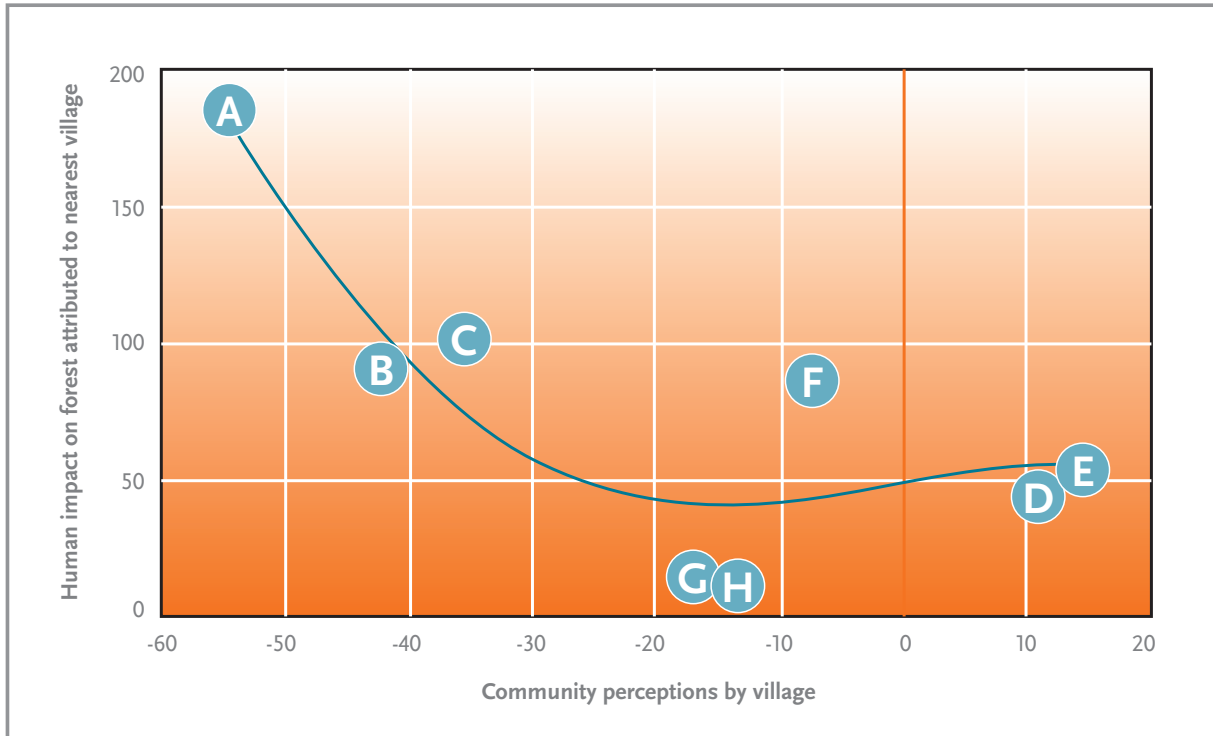

Expanding on the points made in Guidance Note 2 on the link between impact evaluation and routine monitoring and evaluation, this pattern is more productive when it builds on what is sometimes referred to as "closed-loop learning" based on monitoring data. The better the quality of report back and deliberation around monitoring data, the greater the likelihood of effective use of impact evaluations. In other words, if monitoring data is regularly reviewed and deliberated by program stakeholders, then those same stakeholders will be more likely to absorb and act on evaluation findings. Reviews of monitoring data can promote receptivity to impact evaluation by creating a culture of evidence-based deliberation and decision-making.

To put this theme into context, imagine a situation in which ongoing monitoring data are consistently positive and encouraging, but subsequent impact evaluations do not show the expected program effects. Impact evaluation findings of this kind that arrive "out of nowhere" are not likely to realize learning. This highlights the importance of spending time early on integrating the objectives of the evaluation with the logic and established monitoring practices of the program. The more aligned the evaluators are with the ongoing work of implementers, the more likely their findings will be well-received.

Conversely, if the causal logic of the program is clear and measurable from the outset, the learning questions – and therefore the impact evaluation – will be more straightforward.

The figure below shows a comparison between project monitoring data (community harm to an indigenous mixed woodland and corporate-owned hardwood plantation) and community perceptions of the company. In this case, the company seeks to earn profits while providing the community with

Correlation of community impacts on company forests with community perceptions of the company

Y-axis: Human impact on forest attributed to nearest village (0 to 200)
X-axis: Community perceptions by village (-60 to 20)

sustainable forest use benefits. This data is from an actual survey and has been anonymized for this note. It shows a direct relationship between village attitudes towards the company and incidents of harm to the forest caused by that village. The more negative the village views of the company, the more damage the village inflicts on the neighboring forest. This pattern provides the company with a golden opportunity to frame a discussion with the community based on the evaluation findings to explore how it can build a better relationship and reduce harm to the woodlands. Since some villages are far more positive and produce far less harm, there is an opportunity to draw and apply comparative lessons.

Having successfully engaged users in the planning and initial analysis of the evaluation, the last step is to follow through to ensure that the

evaluation messages actually reach their intended users on time and in the form agreed. To increase the likelihood of effective use, it is necessary to do more than thorough consultation before an impact evaluation. Organizations should know and apply the tried and tested craft of knowledge transfer when communicating evaluation results.[7] The first principle is particularly important: the uptake of an idea is dramatically increased when it is understood as a new twist on something that someone already believes.

Finally, to manage the process of communicating with users it is necessary to know if communications are hitting their targets. Smart message

communicators conduct simple after-evaluation report assessments that determine where and how the findings of the evaluation are known. Since this is a factual question and there is little risk of bias in self-collected data, this can be learned by asking the staff who interact with that user to collect this feedback. To provide some quality control, it is also useful to deliver a one- or two-question survey using an online- or cell-phone-based feedback tool. One reliable, open source, and easy-to-use tool is Frontline SMS. The results from these simple feedback exercises should be incorporated into the evaluation scorecards discussed under theme two, below.

To end this theme discussion on a note of realism, we live in a time of very finite resources. It is necessary to set priorities for stakeholder engagement, and to discover the highest value opportunities.

3.   Theme Two: Joined Up Systems and Operations

This theme builds directly on knowing your evaluation purpose and engaging your users. If having a

clear map to users is the starting point, then having the right organizational systems to deliver evaluation findings to those users is the way you get to your destination. This theme was summed up in one interview as "organizational structure" (see text box below). It was the most prominent theme in all the interviews and in the open responses to the online survey, highlighting that evaluations are just beginning to make demands on management systems. Delivering the right findings, in the right form, at the right time puts demands on different organizational systems. Because these demands cut across field operations, advocacy, strategy, communications, and fundraising, taking evaluation seriously will be potentially integrative for organizations.

"Success means that learning from evaluation is integrated into action. It means scaling the learning across the organization, with visible changes. We try to collate a body of evidence from all the impact evaluations that we do so that we can influence decisions at global level."

—Head of evaluation, interview, May 2012

Examples of success offered in the interviews were almost entirely related to meta-analysis across a number of evaluations, and strategic learning with respect to larger themes. One such example cited was CARE's "Strategic Impact Inquiry" method, which was applied to its women's empowerment work.[8] This highlights that evaluation use is more likely when the evaluative effort is able to meet wider strategic organizational objectives, such as learning to improve a cross-cutting theme such as rights-based programming or women's empowerment. According to a CARE staff member playing a leading role in that Strategic Impact Inquiry work, it clarified the value derived from its women's empowerment work and significantly deepened commitment to this approach across the organization.

For large organizations – which were well represented in the survey and interviews – an added layer of complexity was emphasized. A single organization may work in different ways – direct implementation, grant-making, through partners, advocacy, and more – and with a fair amount of decentralization. These circumstances place dramatically different demands on the design, implementation and use of evaluation. Again, however, the integrative potential here is significant. For example, staff involved in direct delivery in the field may be well placed to support an evaluation of partner-based service delivery programs. Staff experience in using evaluations of their work is a strong basis for them to work with partners on how to best use evaluation findings.

> "A lot of what I said about evaluation utilization sums up to organizational structure, including how it impacts communications and decision-making. It certainly is one of the major constraining factors I have faced."
>
> — email received from a head of evaluation after the interview

---

8 Information on the Strategic Impact Inquiry method can be accessed here: http://pqdl.care.org/sii/default.aspx.

The interviewees commonly expressed sensitivity to the burden that doing and using impact evaluation places on staff. They noted that it is important to build evaluation use into existing practices whenever possible, rather than add "new work," which can create negative attitudes towards evaluation and lessen the chances that findings will be carefully considered once the evaluation is complete. Interviewees also pointed out that it is important to designate specific staff with the right skills to communicate evaluation findings, and to make sure that these tasks fit along with other responsibilities.

> An important guideline to reducing the evaluation burden was shared by Mercy Corps: Whenever possible, make full use of existing research data to minimize the burden of original data collection. Mercy Corps illustrates this principle through research it undertook to test the validity of the theories of change of its youth peace-building program in Kenya.[*] The research relied mainly on analysis of data from household surveys that had already been conducted by Mercy Corps and Afrobarometer.[**] By relying on Afrobarometer's published data, original fieldwork took a total of two weeks, and was used primarily to gather qualitative data to triangulate and facilitate the interpretation of the findings from analysis of Afrobarometer's quantitative data. While not "cost-free," Mercy Corps believes the research did better than the typical evaluation in "minimizing the time demands on program personnel."
>
> [*] The report from this research is available at: http://www.mercy-corps.org/resources/youthEDconflictstudy.
> [**] Afrobarometer (http://www.afrobarometer.org/) is an independent research firm that regularly measures the social, political and economic atmosphere in Africa.

Evaluation findings should be included in almost all regular management meetings. One organization, for example, has set evaluation findings or learning questions as a standard agenda item for its team meetings (see box below). Similarly, management discussions of new communications products can include a question or two about how current organizational evaluation findings will be featured. Effective evaluation use requires having a good map of all regular activities where evaluation uses can be integrated relatively painlessly,

and maintaining a record of how faithfully those evaluation-related moments are observed.

The related theme of staff capacity was also frequently cited as a constraint, and the solution offered refers back to the preference to build on existing systems and practices rather than creating new ones. Rather than general and distinct forms of training, which were rated as ineffective in the online survey, one respondent suggested creating "the right spaces and directions for use of findings. We need to provide clear and actionable steps as a result of data."[9]

> Mercy Corps has started a series of meetings called "Evaluation for Organizational Learning Events" (known as EVOLVE). These meetings draw in a wide range of staff working within a common sector, both from the field and headquarters, and engage them in a discussion to reflect on a featured evaluation as well as their own experience in order to answer questions critical to that area of programming. These sector or theme-based evaluation syntheses help to discern key findings from a range of evaluations and summarize important take-away points concerning sector approaches and new program design. While the sessions are recognized as an important source of learning, Mercy Corps leadership emphasize that they require "considerable time on the part of sector support specialists, design, monitoring and evaluation (DM&E) staff, and program team members"- perhaps more time than can be ordinarily sustained. This resurfaces a major theme of this guidance note: the default mode for organizations is to under-invest in evaluation utilization.

Especially because of the trans-organization imperative of evaluation use, appropriate management tools are required. Management consulting firms have a lot of experience with these kinds of challenges. A well-regarded approach is a project scorecard (in this case the project is an evaluation) that aggregates up to higher-level organization scorecards. Indicators should include time and quality of staff inputs as well as after-report metrics of use (such as those introduced in the discussion of theme one, above).

Mercy Corps reports some early success with another common management tool – minimum standards of program management. The standards serve to "link all management functions and processes and sort of 'close the loop' on design to implementation to evaluation to learning and back to design again."

These and other practical suggestions for joined up systems to enable impact evaluation use are echoed in a recently published "Principles into Practice" paper from Millennium Challenge Corporation (MCC).[10] The MCC report helpfully illustrates the particular challenges of using impact evaluation to advance learning in the context of agriculture, where convincing measurement for attribution can be particularly difficult to achieve.

These kinds of solutions require dedicated time (widely seen as less of a problem, though the MCC paper notes a tendency to underestimate the demands of impact evaluation on project implementers) and, more controversially, money. Evaluation is chronically underfunded, which means that *use of evaluation* is virtually unfunded. Several interviewees made the point that evaluation funding tended not to cover the actual costs of creating and implementing a joined-up response. Those who felt that they were making some progress in getting appropriate systems for use in place were doing so with core, unrestricted funding.

Drawing on the discussions of themes one and two, we can summarize this part of the note with an operational checklist for using evaluations.

---

9 "How We Are Using Impact Evaluations": survey report of InterAction members. http://www.interaction.org/document/survey-report-how-we-are-using-impact-evaluations.

10 Katherine Farely, Sarah Lucas, Jack Molyneaux and Kristin Penn, "Impact Evaluations of Agriculture Projects," Millennium Challenge Corporation, Principles into Practice series October 2012. http://www.mcc.gov/documents/reports/paper-2012001116901-principles-impact-evaluations.pdf.

An Operational Checklist for Using Evaluation

Planning & management

1. Develop a value proposition for each potential user.

2. Estimate what evidence will be useful for what user at what time.

3. Recruit a team from across all organization units that will be required to ensure that identified user needs are met.

4. Whenever possible, make full use of existing research data so as to reduce the staff burden of original data collection.

5. Build a communications strategy for evaluations that differentiates internal and external communications and includes user-appropriate reporting formats.

6. Map existing systems and activities for opportunities to tease in small steps for evaluation use.

7. Routinely include discussions of evaluation findings in staff meetings.

8. Track awareness and use through an evaluation scorecard that aggregates up to higher-level organization scorecards. Indicators should include timeliness and quality of staff inputs as well as resulting actions.

User engagement & measuring evaluation use and impact

9. Before the evaluation begins, engage users to test the evaluation hypotheses and proposed indicators, and to determine when and how to best report the findings.

10. Validate tentative findings and deepen interpretations through consultations with users.

11. Conduct assessments one month after the evaluation has been reported to learn where and how the findings of the evaluation are known (awareness) and used.

12. Conduct assessments six months after the evaluation has been reported to learn how the evaluation may have changed users' beliefs and behaviors.

4.   Theme Three: Incentives and Rewards

Given the significant barriers to doing and using impact evaluation, what incentives do people and institutions have to undertake them?

The most persuasive approach to this critical question that we know of comes from organizational learning theorist Edgar Schein. He noted that organizations tend to do things the way they have always done them until the "tried and true" ways no longer work. The reason for this is not obvious. It is, he found, because organizations all have what he calls "learning anxiety" – the anxiety produced by having to shift and learn something new. This anxiety stops organizations from learning. Learning anxiety is so strong that it can only be overcome by "survival anxiety" – the anxiety produced upon realizing that if something does not change, they will not survive. Among international development organizations, survival anxiety is rare, and it is never associated with a failure to use impact evaluations well. If Schein is right, then absent the arrival of new external forces, we cannot expect to see significant progress on evaluation use.

New external forces are inexorably raising survival anxiety. Two relate to practice trends: greater advocacy and more professional management.

As more organizations take on advocacy activities, to be successful (in Schein's terms, "to survive") they need to support their positions with valid evidence.

The general trend toward more formal and professional management means more use of metrics. More metrics creates more demand for robust impact evidence.

4.1. A growing climate of accountability

The major external force, however, is the growing climate of accountability over public benefits in general – governmental and private/nongovernmental – and particularly transparency-based approaches to accountability. As austerity deepens, so too does the pressure to provide better evidence of results.

The U.S. Office of Management and Budget issued new standards in 2012 that promote programs with rigorous studies of effectiveness.[11] Federally funded programs in human services, education and health are required to meet increasingly stringent impact measurement standards. While the evaluation policy at USAID is less stringent, it is moving in the same direction and requires USAID to publish all evaluations it funds.[12]

In what may be a telling direction for accountability demands on U.S.-based international development organizations going forward, the 2012 Foreign Appropriations bill emerged from Senate-House conference with a provision:

> "*to enhance the effectiveness of humanitarian aid through a continuous and accurate flow of feedback data, obtained independently. Such feedback will ensure that both aid agencies and donors have a clearer understanding of the perceptions of affected populations, whether the aid they receive is relevant to their needs, how much they trust the people helping them, and whether they find it worthwhile to provide feedback. When reported to Congress*

---

11 "Use of Evidence and Evaluation in the 2014 Budget," Office of Management and Budget. http://www.whitehouse.gov/sites/default/files/omb/memoranda/2012/m-12-14.pdf.

12 http://blog.usaid.gov/2011/01/usaids-evaluation-policy-setting-the-standard/.

and published, feedback data of this kind, which provides the basis to compare and contrast the perceived performance of aid providers, may offer an incentive to turn the principle of accountability to beneficiaries into practical improvements on the ground and cost savings."

Going even further than setting higher standards of evidence reporting, many governments around the world are now seriously looking at structuring their funding so that they only "pay for success," as defined by achieving measurable pre-identified outcomes.[13] This approach has prompted a financing innovation known as the social impact bond, which leverages private sector investment and whose returns are paid for by governments when, and only when, outcome targets are met.[14]

4.2. Internet-powered change in the model of giving for international development

Private nongovernmental accountability demands are also growing rapidly, but here the major force driving change is the Internet, which comes with an existential threat to "business as usual" for international NGOs. Internet-based giving marketplaces such as GlobalGiving and Kiva offer a new model of giving that enables individual citizens to give directly to grassroots groups in the developing world via their giving marketplace platforms. The Hewlett Foundation has compiled a database of over 100 of these "giving platforms."[15]

Giving direct to local groups via an Internet marketplace carries an alternative model of accountability. Giving platforms assemble evidence of impact by the organizations listed on their sites, and inform givers about how to allocate their giving to the "most effective" organizations, often in partnership with third-party charity raters.

The largest charity rater in the world, Charity Navigator, which now gets over 3.5 million discrete users a year, has found that its ratings are influencing billions of dollars a year in individual giving. Now 10 years old, Charity Navigator until recently used a rating model exclusively focused on financial health and good governance. Partly in response to criticism that this two-dimensional approach does not equate with charity effectiveness, Charity Navigator has announced that it will begin to rate charities on the quality of the reported evidence of their results.[16]

4.3. Cultivating a learning culture

With these winds of change filling their sails, evaluation "intrapreneurs" can do a lot to stay ahead of the accountability curve.

Impact evaluation use fits within an organization's culture of learning. Job one is to create a strong and enabling culture of learning in the organization. There is a large literature on this, including specifically relating to international NGOs and learning. These six steps are validated in the literature:

1. **Leadership buy-in**. Learning cultures are made, not born. They require strong and consistent support from senior leadership. If starting from a low base, the M&E team can work with senior management to deliver a

---

13 The U.S. government is among the leaders. See, for example, Office of Management and Budget, "Paying for Success." http://www.whitehouse.gov/omb/factsheet/paying-for-success. This approach is fast spreading across the world.

14 For a recent McKinsey and Co. report on the potential applicability of social impact bonds, see http://mckinseyonsociety.com/social-impact-bonds/.

15 This database is an internal working document for a group of approximately 35 US nonprofit sector leaders who have come together under the label "Markets for Good." It can be found at http://www.marketsforgood.org.

16 "Where we are Headed," Charity Navigator. http://www.charitynavigator.org/index.cfm?bay=content.view&cpid=1193.

series of messages about the importance of, and new practices to make space for, learning from impact evaluations.

2. **Conduct an internal review of current incentives and disincentives for learning**. Once you know the ways in which learning is discouraged (for example, in how failure is perceived and penalized), it is possible to restart a culture-building process with a bonfire of learning disincentives.

3. **Create opportunities for reflection at which learning questions are discussed**. These can and should be diverse and overlapping in nature, ranging from:
   a. "Meet the evaluator" brown bag lunches (which work best if they start at the design stage of an evaluation and continue through to discussions of findings); to
   b. Formal reviews of evaluation progress at management team meetings; to
   c. Short presentations of results at a staff meetings; to
   d. Presentations at regular programmatic staff meetings that build a plan to ensure that evaluation results are utilized.

4. **Start giving rewards for learning achievements** – as opposed to "coming out well" in an evaluation – in the form of recognition, and even remuneration.

5. **Build the use of evaluation into job descriptions and performance appraisals**. At a minimum, anyone who commissions or contributes to an evaluation should be held accountable to see that findings are utilized.

   a. It is important to recognize that with personal accountability comes another kind

of anxiety: "Will I be judged to be performing badly?" Incorporating accountability for evaluation use into job descriptions must be about creating the conditions for frank and open discussions of results. Accountability is for enabling learning, not for getting "good" evaluation results. The greatest failure in development is the failure to learn from our work.

6. **Deploy and manage vigorously the evaluation utilization tools described in this note under theme two**.

4.4. Collective learning

In line with Schein's insights regarding the epic struggle between learning anxiety and survival anxiety, the single most important thing that the international development sector can do to encourage effective use of impact evaluations would be to launch an impact evaluation use benchmarking project. Because it would be inexpensive to run, it could be funded by nominal subscription of users, whose participation would be voluntary.

Here are six steps to establish such a benchmarking project:

1. Recruit an initial group of at least 10 organizations willing to participate.

2. The founder group sets out the fees and rules of engagement for the project around confidentiality and transparency and the inclusion of additional participants in the first year.

3. The founder group sets out a set of no more than 12 indicators of evaluation use to be tracked and reported by all participants.

4. Participants track and report on agreed measures on a quarterly basis.

5. Reported measures are analyzed and reported in line with agreed rules of engagement.

6. Participants meet in person and virtually halfway between quarterly reports to interpret the data and exchange experiences.

After four reporting cycles, conduct a formative evaluation to revise metrics and learning procedures and develop a plan to take the project forward, including recruiting more members.

One can begin to see what may be precursors to such a benchmarking project in specific sectors. For example, in the agriculture sector an informal working group of government aid agencies and international foundations has created a Food Security Learning Agenda that sets out high priority questions

> *"...for which evidence and answers are lacking and which could contribute the most to efforts to improve food security programming around the world. The Food Security Learning Agenda represents a concentrated effort to promote evidence-based learning among partners working in the agriculture, economic growth, and nutrition sectors and provides a framework of priorities toward which partners can focus and align the programming of their M&E efforts."*[17]

Summary

The InterAction impact evaluation guidance note series provides a complete, introductory reference set for international development organizations.

---

[17] Unpublished internal working draft dated 20 September 2012.

It sets out the basic why's and how to's of impact evaluation, while providing citations to more detailed and technical literature.

This note, the last in the series, focuses on the use of impact evaluations. The core argument of this note is that the true gold standard of impact evaluation is not one or another methodology, but whether and how well an evaluation is used to enhance and accelerate development outcomes.[18] It should be clear to the readers of this guidance note series that the notes' authors do not view any one evaluation method as inherently superior. Rather, they have tried to provide the concepts and practical guidelines that will enable practitioners to design and conduct the impact evaluations best suited to their purposes. And whatever evaluation methodologies are employed, this final guidance note emphasizes that the most important question that you need to ask and answer with respect to evaluation is, "So what?"

The determinative measure of an evaluation is how that evaluation informs subsequent action. The recommended approach and practices set out in this guidance note are offered to help ensure all your impact evaluations excel against this measure. The cautionary alarm sounded throughout this note is that, absent well-laid plans supported by a range of management practices, organizations default to very weak use of evaluation findings. Even the best plans and practices are unlikely to sustain rigorous evaluation use over time unless internal and external incentives reinforce effective use of evaluations. The note argues that current trends favor more and better use of evidence. But the balance of incentives still strongly favors the *status quo ante*, in which evaluation use is not carefully tracked and managed.

---

[18] The author acknowledges Jodi Nelson, Director of Impact Planning and Improvement at the Bill & Melinda Gates Foundation, for this formulation of the evaluation gold standard.

References

Interviewees

Names, titles and organizational affiliation at the time of interview

- Amy Biel, Monitoring and Evaluation Specialist, Heart to Heart
- Anna Young, Senior Director, Strategy and Learning, Mercy Corps
- Dale Hill, Senior Monitoring and Evaluation Advisor, American Red Cross
- Elena Vinogradova, Research Scientist, Education Development Center
- Guy Sharrock, Senior Advisor for Monitoring & Evaluation, Catholic Relief Services
- Holta Trandafili, Design, Monitoring & Evaluation Specialist, World Vision
- Jeannie Annan, Director of Research, Evaluation and Learning, International Rescue Committee
- Kristin Penn, Senior Director of Agriculture, Millennium Challenge Corporation
- Maby Palmisano, Senior M&E Director, ACDI/VOCA
- Maliha Khan, Director, Learning, Evaluation and Accountability Department, Oxfam America
- Rienzzie Kern, Senior Director Planning, Monitoring & Evaluation, Heifer International

Published materials

Bamberger, M. and Kirk. A. *Making Smart Policy: Using Impact Evaluation for Policymaking*. World Bank, 2009. http://siteresources.worldbank.org/INTISPMA/Resources/383704-1146752240884/Doing_ie_series_14.pdf.

Cousins, J. and Leithwood, K. "Current Empirical Research on Evaluation Utilization," *Review of Educational Research*, 1986. Vol. 56, No. 3, Pp. 331-364.

Farley, K. et al. "Principles into Practice: Impact Evaluations of Agriculture Projects," Millennium Challenge Corporation, 2012. http://www.mcc.gov/documents/reports/paper-2012001116901-principles-impact-evaluations.pdf.

Hallam, A. "Harnessing the Power of Evaluation in Humanitarian Action: An initiative to improve understanding and use of evaluation," ALNAP, 2011. http://www.alnap.org/pool/files/evaluation-alnap-working-paper.pdf.

Johnson, K. et al. "Research on Evaluation Use: A Review of the Empirical Literature From 1986 to 2005," *American Journal of Evaluation*, 2009. Vol. 30 No. 3, Pp. 377-410.

Ledermann, S. "Exploring the Necessary Conditions for Evaluation Use in Program Change," *American Journal of Evaluation*. 2012. Vol.33 No. 2, Pp. 159-178.

Patton, M. *Utilization-Focused Evaluation*, 4th edition. London: SAGE, 2008.

"Actionable Measurement at the Gates Foundation," The Bridgespan Group Blog. http://www.bridgespan.org/Blogs/Measuring-to-Improve/September-2012/Actionable-Measurement-at-the-Gates-Foundation.aspx#.UHvzYRi2I4A.

"As I Recall—Or How to Take Advantage of Less-Than-Successful Evaluation Experiences," *The Canadian Journal of Program Evaluation Special issue* 2010. Vol. 25 No. 3. http://cjpe.ca/site.cgi?s=4&ss=21&_lang=en&volume=2010//3.

*Presents 11 cases of evaluations that failed to pro-
duce the expected results.*

"Designing a Better Partnership," Center for
Effective Philanthropy Blog. http://www.effective-
philanthropy.org/blog/2012/10/designing-a-bet-
ter-partnership/?utm_source=feedburner&utm_
medium=email&utm_campaign=Feed%3A+effec
tivephilanthropy+%28The+CEP+Blog%29.

"Making an Impact: Impact measurement among
charities and social enterprises in the UK," New
Philanthropy Capital. http://www.thinknpc.org/
publications/making-an-impact/.

"State of Evaluation: Evaluation Practice and
Capacity in the Nonprofit Sector," Innovation
Network. http://www.innonet.org/client_docs/
innonet-state-of-evaluation-2010.pdf.

"How We Are Using Impact Evaluations," Keystone
Accountability. http://www.interaction.org/docu-
ment/survey-report-how-we-are-using-impact-
evaluations.