PRACTICAL CHALLENGES OF RIGOROUS IMPACT EVALUATION IN INTERNATIONAL GOVERNANCE NGOs: Experiences and Lessons from The Asia Foundation

Nina Merchant-Vega, Assistant Director, Economic Development Programs, The Asia Foundation
PRACTICAL CHALLENGES OF RIGOROUS IMPACT EVALUATION IN INTERNATIONAL GOVERNANCE NGOs: Experiences and Lessons from The Asia Foundation

Nina Merchant-Vega, Assistant Director, Economic Development Programs, The Asia Foundation
The Asia Foundation is a nonprofit, non-governmental organization committed to the development of a peaceful, prosperous, just, and open Asia-Pacific region. The Foundation supports Asian initiatives to improve governance, law, and civil society; women’s empowerment; economic reform and development; sustainable development and the environment; and international relations. Drawing on nearly 60 years of experience in Asia, the Foundation collaborates with private and public partners to support leadership and institutional development, exchanges, and policy research.

www.asiafoundation.org
TABLE OF CONTENTS

1. INTRODUCTION ........................................................................................................................................... 1

2. WHAT IS RIGOROUS IMPACT EVALUATION? ................................................................................................ 2

3. RIGOROUS IMPACT EVALUATION CHALLENGES ......................................................................................... 5

4. WHAT CAN BE DONE ..................................................................................................................................... 9

5. CONCLUSIONS .............................................................................................................................................. 13

REFERENCES ...................................................................................................................................................... 14
1. INTRODUCTION

Over the last several years, nongovernmental organizations (NGOs), development practitioners, and academics have become increasingly interested in rigorous impact evaluation. Proponents of impact evaluation argue that having clear, measurable evidence of program outcomes will help practitioners to better understand how to design and implement effective policies as well as help policymakers to enhance overall aid effectiveness.

As a result, the argument for impact evaluation has gained traction among major donors. Over the past several years, donors like the U.S. Agency for International Development (USAID), the Australian Agency for International Development (AusAid), the UK Department For International Development (DFID), the World Bank, the International Finance Corporation (IFC), the German Agency for International Development (GIZ), the French Development Agency (AFD), and the Canadian International Development Agency CIDA have made efforts to incorporate impact evaluations in their work. As a natural extension of this, implementing partners that work with these agencies have also become increasingly interested in impact evaluation and have stepped up efforts to incorporate such evaluations into their own work.

However, despite the enthusiasm of academics, donors, and practitioners, the fact remains that rigorous evaluations are still rare (CGD, 2006). This paper aims to shed light on the institutional and programmatic challenges involved in implementing rigorous impact evaluation through the experience of The Asia Foundation (the Foundation), an international NGO, focused primarily on governance and governance-related issues.1 The Foundation is a relatively large organization, with substantial program resources and significant reach and history in Asia. Therefore, understanding the impact of its programs is of paramount importance not only to its donors, but program managers and staff. However, as in the case of other international NGOs, very few rigorous impact evaluations have been planned because of barriers related to how programs are funded and other challenges.

---

1 Founded in 1954 The Foundation supports Asian initiatives to improve governance and law, economic development, women's empowerment, the environment, and regional cooperation. Drawing on nearly 60 years of experience in Asia, the Foundation collaborates with private and public partners to support leadership and institutional development, exchanges, and policy research. With 17 offices throughout Asia, an office in Washington, DC, and its headquarters in San Francisco, the Foundation addresses these issues on both a country and regional level. In 2010, the Foundation provided more than $98 million in program support and distributed nearly one million books and journals valued at over $42 million. The Foundation is funded primarily by bilateral and multilateral donor agencies including USAID, AusAid, DFID, DANIDA, CIDA, the World Bank, IFC, and ADB among others. In 2010, total funding for programs was more than $130 million.
2. WHAT IS RIGOROUS IMPACT EVALUATION?

Rigorous impact evaluations aim to answer the central question imperative to any development program: “how would outcomes have been different if the intervention had not been undertaken?” While this question seems unremarkable to those outside the development community, donors and development professionals know how difficult it is to answer. It is essentially a cause-and-effect question that deals with figuring out what the outcomes would be both with and without the intervention rather than just before and after an intervention. This distinction is crucial, as there are many intervening factors that can affect the outcome we observe outside of the intervention. If we simply look at the outcomes before and after an intervention, we might falsely attribute an outcome to an intervention that had no real impact.

This focus on attribution is the defining feature of rigorous impact evaluation. In order to accurately attribute an outcome to a certain intervention, an evaluator must estimate a “counterfactual” – i.e. what the outcome would have been for program recipients if they had not participated. Thus, in practice, impact evaluation requires an evaluator to have a control group that is comparable to the recipient (treatment) group (Gertler et al., 2010).

In this way, rigorous impact evaluation stands in contrast to many of the standard monitoring and evaluation (M&E) tools used in development which attempt to measure program processes and targets (are we implementing the program on time?) or outputs (how many people trained, how many seminars conducted?) or even short-term outcomes (are more people participating at local council meetings, etc?). While these standard tools are useful—and indeed necessary for development practitioners—they often fall short of clearly identifying program impact.

2.2 Types of Rigorous Impact Evaluation

In recent years, rigorous impact evaluation practice has become conflated with experimental evaluations (i.e. randomized controlled trials). While experimental evaluations are considered the “gold standard” for evaluating the effectiveness of development programs—and some academic development economists consider them the only kind of evaluations that meet the standard of rigor—most practitioners (and institutions that support practitioners) have a more expansive view of what “rigor” means. For example, International Initiative for Impact Evaluation² (3IE) defines rigorous evaluation as:

“Analyses that measure the net change in outcomes for a particular group of people that can be attributed to a specific program using the best methodology

---

² 3IE funds quality studies that will have a real policy impact and affect many lives. In terms of standards, this means only studies that are built around a credible counterfactual with an evaluation design based on the underlying program theory to learn what works and why, and also at what cost.
available, feasible and appropriate to the evaluation question that is being investigated and to the specific context.\textsuperscript{3}

Thus, the method for a rigorous evaluation is not necessarily proscribed, but rather it must meet certain criteria while being both feasible and appropriate to the program and context. In a practical sense this translates into the following guidelines for inclusion in the 3IE database:

1. Impact evaluations must address the attribution question to establish cause and effect between programs and both intended/unintended outcomes;
2. Impact evaluations should have clearly defined participants and non-participants (treatment and control group) and a sufficiently large sample size to make statistical analysis of sufficient power feasible;
3. The control group must be selected in such a way as to a) minimize the chances of spillover effects and contamination; b) allow for confounding factors; and c) remove possible selection bias;
4. Impact evaluations must employ either the experimental or quasi-experimental approaches.\textsuperscript{4}

Under 3IE’s definition, both quasi-experimental and experimental approaches are considered rigorous. Related organizations such as the Independent Evaluation Group of the World Bank and The Network of Networks on Impact Evaluation (NONIE)—a group comprised of the OECD’s Development Assistance Committee (OECD/DAC), Evaluation Network, and the United Nations Evaluation Group (UNEG) among others—have taken similar stances on what evaluations qualify as rigorous.

3IE and others also encourage a mixed method approach as a way to both flesh out the results of the evaluation and ensure policy relevance. While no institution involved in rigorous impact evaluation advocates for a purely qualitative approach, most maintain that qualitative work, in addition to quantitative work, can help in the interpretation of data patterns.\textsuperscript{5}

Using these definitions, impact evaluation methods can be broadly categorized into experimental and quasi experimental categories (Table 1, next page).

\textsuperscript{4} Ibid. (2011).
\textsuperscript{5} Ibid. (2011).
**Table 1: Basic Impact Evaluation Methods**

<table>
<thead>
<tr>
<th></th>
<th>Experimental</th>
<th>Quasi Experimental</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Randomized Controlled Trials</td>
<td>Difference-in-Differences</td>
</tr>
<tr>
<td>Random Assignment of Control &amp; Treatment Groups</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Prospective or Retrospective</td>
<td>Prospective</td>
<td>Retrospective</td>
</tr>
<tr>
<td>Complexity of Statistical analysis</td>
<td>Low-High*</td>
<td>High</td>
</tr>
<tr>
<td>Strong Assumptions of Comparability</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

- **If randomization is done perfectly, without complicating factors, the complexity of the statistical analysis will be relatively low. However, complicating factors like spillovers or non-compliance might add to the complexity.**
- **While close to cut-off control and treatment groups may look very similar, similarities diminish the farther one gets away from the eligibility cut-off.**

In an experimental approach, an intervention is randomly assigned to some units (individuals, communities, schools), while others form a control group prospectively. Researchers then compare the averages of the two along some set of defined outcomes. Proponents of the experimental approach argue that this method allows researchers to cleanly estimate casual impacts by removing selection bias, thus eliminating the need for evaluators to make assumptions about the comparability between the two groups (Cohen & Easterly, 2009). Moreover, if the randomization is done correctly, the complexity of the statistical analysis is often low. In this way, researchers, policy-makers and development practitioners have a clear conception of whether the program works as well as precise measure of the size of the impact.

Quasi-experimental approaches use fairly sophisticated statistical analysis to estimate outcomes between non-randomly assigned control and treatment groups. The most commonly used quasi-experimental methods are difference-in-differences, matching, and regression discontinuity.

Difference-in-differences compares changes in outcomes overtime between control and treatment groups; matching techniques rely on observable characteristics to construct a treatment group. Both techniques are implemented retrospectively. While retrospective implementation may provide more flexibility for evaluators, in order to accurately establish attribution, it also requires evaluators to make strong assumptions of comparability between the control and treatment groups. This may shed doubt on the validity of the results (Gertler et al., 2010).
Regression discontinuity uses a program eligibility cut-off—such as age or income—to create control and treatment groups. For example, a poverty alleviation program may have an income cut-off in which those that just qualify for the program are compared with those that fall just outside of the program. While very close to the cut-off score control and treatment groups tend to be very comparable, farther away from the cut-off, groups may not be comparable at all. Thus, results can only really be valid for those who are very close to the cut off score (Gertler et al. 2010). In addition, like the other quasi-experimental methods, regression discontinuity is generally done retrospectively and usually requires sophisticated statistical analysis.

Undoubtedly, experimental, difference-in-differences, matching, and regression discontinuity each carry their own set of unique advantages and disadvantages as methods. While differing from each other in specific use and implementation, each method fits within the broadly accepted definition of impact evaluation.

3. RIGOROUS IMPACT EVALUATION CHALLENGES

Recent efforts to incorporate more rigorous impact evaluation design into the programs of donors, NGOs, and other development practitioners have been spearheaded by international organizations such as the International Initiative for Impact Evaluation (3IE), and the Center for Global Development (CGD). Highly-regarded academic economists such as Esther Duflo and Abhijit Banerjee at MIT’s Abdul Latif Jameel Poverty Action Lab (J-PAL), Dean Karlan of Innovations for Poverty Action (IPA), and Ted Miguel of the Center for Evaluation for Global Action (CEGA) have also been vocal proponents of rigorous impact evaluations.

Impact evaluations have garnered attention in the development community because of their promise to identify successful and cost-effective development interventions through evidence rather than “fad” theories or political paradigms. Pritchett (2002) argues that many bilateral and multilateral donors are staffed with “true believers,” professionals who are attached to a particular paradigm or approach to solving problems, without much evidence that those paradigms or approaches work. He wrote of World Bank employees:

“The World Bank is for the most part staffed by internationally recruited, high quality, highly trained, well-meaning, and experienced professionals and is arguably the premiere development institution. And yet nearly all World Bank discussions of policies or project design had the character of ‘ignorant armies clashing by night’—there was heated debate amongst advocates of various activities but very rarely any firm evidence presented and considered about the likely impact of the proposed actions.”

Rigorous impact evaluation has caused so much excitement precisely because it can provide clear evidence free of ideological or political bias. In development work this

---

represents a radical break from the way programs were (and still are) usually designed in which approaches are often taken as axiomatic (“participation will strengthen democracy,” “anti-corruption campaigns will reduce corruption.” etc.).

The argument for impact evaluation has gained traction among government’s major donor agencies—at least on paper. For example, President Obama’s 2010 UN speech indicated his administration’s commitment to evidence-based policy, stating, “guided by the evidence, we’ll invest in programs that work and end those that don’t. Because we need to be big-hearted and hard-headed.” In June 2010, Ruth Levine, then Senior Advisor for Evaluation at USAID articulated the specifics of this policy, indicating that USAID was working on a plan to set priorities for rigorous evaluations. That same year, AusAid formed a partnership with 3IE, which specializes in funding large-scale randomized trials program evaluations and other rigorous evaluations. These endeavors are just some of the most recent examples in a string of efforts over the past several years to incorporate rigorous evaluations into the programs sponsored by donors like USAID, AusAid, DFID, the World Bank, IFC, AFD, GIZ, CIDA and others.

While these efforts clearly illustrate a support of rigorous impact evaluation, the fact remains that these types of evaluations continue to be uncommon in the development community. Pritchett (2002) and CGD (2006) describe why impact evaluation continues to be a very rare tool. Much of the problem has to do with the institutional incentives; donors, NGOs, and other development practitioners, risk diminished budgets, staff and prestige if their programs are shown to be ineffective. Blattman (2008) also describes how experimental evaluation, while useful, might not always meet the needs of practitioners. This is particularly true for those involved in more complex types of programming where understanding causal mechanisms is more important than fully quantifying causality and return on investment.

The experience of The Asia Foundation points to such institutional and programmatic incentives as the main constraints to implementing rigorous impact evaluations. While the Foundation has made concerted efforts to increase staff capacity to implement impact evaluations and incorporate more rigorous evaluations into its programming, to date, very few rigorous experimental evaluations have been supported or planned because of substantial institutional disincentives and practical programmatic limitations.

3.1 Institutional Challenges

First and foremost, the Foundation has faced institutional challenges related to “demand, money and incentives” (CGD, 2006). Despite donor rhetoric, the fact remains that very few donors demand or fund any kind of rigorous evaluation. This partially has to do with

---

Donor disincentives in investing in knowledge generation, which as Levine (2005) points out is a public good, and like all public goods, tend to be underinvested in. Compounding this problem is the fact that investment in evaluations involves an inter-temporal trade-off. The investment in the evaluation, particularly randomized controlled trials, must be made at the start of the program; however benefits from the knowledge generated will not be realized until the end of the project. In this way, the benefits will flow to the next round of projects, but may not have a benefit to the current project, so the investment is not made.

Donor agencies also face powerful political economy disincentives against impact evaluation investment. As Pritchett (2002) points out (and as mentioned above), many donor institutions are wedded to a particular development paradigm or issue, even if there is little evidence to suggest these paradigms work or the issue is of great importance. Since the outcome of an evaluation is not guaranteed to be positive at the outset, these institutions are better off not investing in an evaluation that might produce a negative outcome. This is further complicated by bureaucratic inertia. It is well-known that cutting even the most ineffective programs is difficult when faced with well-organized, bureaucratic stakeholders that want to continue along a certain paradigm.

Moreover, donor agencies are subject to political oversight by elected officials, and in many cases are considered overt tools of foreign policy. As a result, they must carry out programs associated with the economic and foreign policy agendas of the governments they work under. Negative or unfavorable evaluation results may threaten such agendas from a political perspective. These agencies are also under incredible pressure to show that every program and initiative they undertake is effective because they are accountable to lawmakers and, ultimately, taxpayers. Thus, institutions may risk their budgets, staff and programs by agreeing to rigorously evaluate projects.

These problems ultimately trickle down to implementing partners like The Asia Foundation. At the most basic level, if donors are unwilling to ask for and fund impact evaluations, it is difficult for implementing partners to do them independently. And, like donors, NGOs face upfront investment risk. At the Foundation, program managers are often interested in conducting rigorous impact evaluations, but simply cannot commit significant program resources to evaluation because that would decrease the pool of resources to the program itself. For example, the Foundation’s offices in both Sri Lanka and Bangladesh have been interested at various points in using rigorous impact evaluation to better understand the impacts of community policing programs. However, donors have not provided funds and support for such evaluations, meaning that these offices have had to settle for alternative methods of evaluation.

Lastly, even in a hypothetical case in which a program manager could divert some program resources to conduct an evaluation, undertaking an impact evaluation without the full support of the donor not only exposes program managers to the inter-temporal risks mentioned above, but also to the risk that donors will simply be displeased that resources were not used for the program itself.
3.2 Programmatic Challenges

In addition to the considerable institutional disincentives, there are also significant practical programmatic challenges to conducting experimental evaluations for The Asia Foundation. First, as Ravallion (2009) aptly states in the case of experimental evaluations, “plainly randomization is only feasible to a non-random sub-set of interventions and settings relevant to development.” This is also true in the case of quasi-experimental impact evaluation. For example, it is impossible to use accepted impact evaluation techniques to evaluate programs which have national impacts because it’s obviously impossible to have a “control Vietnam.” This is a problem for a governance-focused NGO like the Foundation, which is interested in governance reform efforts at the national level, as well as the sub-national levels. Such reforms might be central to a country’s overall development, but impossible to evaluate using impact evaluation techniques.

However, even if we leave aside these types of programs and look only to a set of programs which could be evaluated with clean control and treatment groups, the issues of targeting and level of evaluation often complicate the matter. Donors often mandate certain aspects of programs, such as the areas in which the program must be implemented or the populations that must be targeted. This makes it impossible to randomize roll-out of a program, and in some cases even to conduct rigorous quasi-experimental evaluations. For example, in Sri Lanka, many Asia Foundation projects are concentrated in the northern and eastern provinces, which are emerging from many years of war. However, because the Foundation’s projects focus on governance and development, many interventions are clustered at the district or province-level rather than at individuals or small groups of people. Since there are so few districts and only two provinces, it is nearly impossible to have enough units to divide into clean control and treatment groups (Gertler et al., 2010). And in the case of a place like Sri Lanka, it would be inappropriate to use non-program districts like the southern or western provinces as comparison groups because their histories and ethnic compositions are so radically different than the northern and eastern provinces. This further reduces the pool of impact evaluations that NGOs like the Foundation are likely to undertake.

These problems are further compounded by limited staff capacity in the field. Because it is possible to partner with qualified research organizations or consultants, it is not necessary for program managers to understand how to conduct complex impact evaluations on their own; however it is necessary for them to have a basic understanding of how impact evaluation works. This knowledge can help them work with researchers to identify and design evaluations. Such knowledge is lacking in NGOs like the Foundation, which have excellent program implementers, but fewer people with the quantitative social science research skills to carry out such evaluations. In March 2010, the Foundation hosted an internal workshop with CEGA for its own staff to strengthen their understanding of impact evaluation. However, more would need to be done to build capacity among implementers.
Lastly, in terms of pure program learning, the Foundation and other governance NGOs, often ask: do impact evaluations always tell us what we want to know in this program context? As Blattman (2008) observes, even if a perfectly designed impact evaluation can tell us the impact in quantitative terms, “we are often in no position to say specifically who or what is responsible for a positive or negative impact.” For governance programs that generally have several components and that are often conditional on context and process, this is a very real concern that feeds back into concerns about institutional political economy risk. For example, the election of a new government might completely derail a program that would have been effective under the previous government because the new locally-elected mayors have no intention of implementing a program that the previous government agreed to. In this case, the most important variable in the potential success of the project was local commitment to implementation, but an impact evaluation would not tell us that. An impact evaluation could only tell us if the program succeeded or failed. In this way, governance program managers often see impact evaluation as an inherently limited tool that might not give them the most valuable information they need on how a program works.

Program managers at the Foundation and other governance NGOs are often skeptical of how much the results of impact evaluations can really impact policymaking in the face of political constraints. As Romer (2009) points out, “policy makers are constrained, but rarely ignorant.” While rigorous evaluations have the potential to help push along some reforms by providing credible evidence to policymakers, this is certainly not always the case. For example, in Colombia, a private-school voucher program failed to be adopted after a successful experimental evaluation (Angrist et al., 2002). This is because policymaking is ultimately a political game, not a technical one. Thus, in many instances, program managers are reluctant to invest in impact evaluations that may not help the reform agenda despite the evidence.

4. WHAT CAN BE DONE?

To-date, the Foundation has made some investments to try to enhance staff capacity and focus the institution more on evaluation. As noted, the Foundation has been pro-active in trying to enhance the capacity of its staff. The Foundation also created an internal Evaluation Task Force in 2007 to support evaluation initiatives and shape its evaluation policy. In addition, the Foundation is trying to build relationships with independent evaluation donors like the International Initiative for Impact Evaluation (3IE) to find independent sources of funding for impact evaluations. These are small, but potentially significant investments that aim to elicit ideas and debate the issues surrounding impact evaluation.

However, if experimental evaluation is to increase at the Foundation and other governance NGOs in a significant way, institutional changes need to take place at both the donor and implementing partner-levels. These institutional changes need to create better incentives for organizations to undertake such evaluations. In addition, serious attention needs to be paid to the programmatic concerns of implementers. If evaluations
don’t serve the needs of implementers as well as other stakeholders, they will rarely be undertaken voluntarily by NGOs and other implementing partners.

4.1 Institutional Investments for Evaluation

In the short-term, it will be difficult to change the structural incentives of donors and implementing partners. Ultimately, implementing partners will be accountable to donors; donors will be accountable to elected politicians who will place pressure on the donors to prove that their projects are achieving results. However, there are clear steps that can be taken to enhance institutional commitment to evaluation even within this flawed incentive structure.

Individually, bilateral and multilateral donors can begin investing in a serious way in rigorous evaluation with staff and resources dedicated solely to the task of creating good evaluation policy. For example, in 2010 USAID consolidated the policy planning and evaluation functions, which were previously spread over various parts of the organization, into the Bureau of Policy, Planning and Learning (PPL). Such consolidation is designed to help focus the institution to develop an internally consistent evaluation policy and pool of resources. This will ensure that individual parts of the USAID bureaucracy are not taking an ad-hoc approach to rigorous evaluation and that resources are available for evaluations. The PPL is also responsible for putting together a training program to build a core set of evaluation skills within the agency to ensure that staff understands the various types of evaluation methods and tools and speaks the same language across the institution.

USAID’s recent 2011 evaluation policy also sets priorities for how to map methods to different types of questions. As the former Deputy Assistant Administrator for the PPL, Ruth Levine stated at the State Department Conference on Program Evaluation in 2010:

“Yes, there are questions about the ultimate impact of programs. There are also questions about the implementation of programs, better and not as successful ways to implement programs. And those are equally and sometimes more important that those around the ultimate impact, and they too require strong methods.”

Indeed, the policy statement takes into account the programmatic needs of implementing partners, focuses resources on performance learning—normative and descriptive questions about what a program has achieved and how it is being implemented—as well as impact evaluation (USAID, 2011). In this way, the policy does not take a “one-size fits all” approach and demand rigorous evaluations of all programs, but instead tries to understand what type of evaluation is needed for any given program. Such nuance is important for reducing the perceived institutional and programmatic risks implementing partners face.

---

In addition, the policy sets clear parameters for funding and guidelines. USAID has committed to devoting approximately 3 percent of total program dollars on average to external performance, including impact evaluations, and acknowledges that in some cases this may require the reallocation of resources away from project implementation, to evaluation and knowledge generation (USAID, 2011). In recognizing this trade-off USAID is clearly signaling the importance it is placing on evaluation and learning, which gives implementing partners a better sense of how to weigh resource allocation between project implementation and impact evaluation.

Other donor agencies have also stepped up efforts. In 2010 DFID announced that it would create an Independent Commission for Aid Impact (ICAI), separate from DFID that will incorporate experimental evaluation design among other evaluation tools. By existing outside of the current DFID bureaucracy, ICAI will not only help shape policy and create a pool of resources, but also bypass some of the more difficult incentive problems associated with internal evaluation. Since ICAI will be accountable for conducting evaluations and monitoring aid spending—rather than doing both of these tasks in addition to implementing programs—there will be less institutional conflict of interest.

Efforts such as these represent the most serious push for evaluation reform within aid agencies in recent memory. They focus squarely on coordinating evaluation policy, dedicating resources to evaluation, improving standards for evidence, and building capacity within institutions. These efforts, if they succeed, will help create better demand and incentives for experimental evaluations.

If such donor efforts expand support for rigorous evaluation, implementing partners must also step up their own investments to be able to carry them out. The first investment implementing partners must make is in building capacity. It is important to invest in learning for existing staff, particularly field-based staff, who intimately understand the challenges of program implementation and who will be called upon to help design evaluations, and in some cases, work with external evaluators. Capacity can also be built by hiring-in new staff that are trained in evaluation. Such capacity can help ameliorate “fear of the unknown” and help managers better understand what is possible in terms of rigorous evaluation.

As noted, in March 2010, the Foundation took the first step toward building this capacity by partnering with CEGA to hold a training workshop on impact evaluation for program managers and field staff. The workshop focused on helping staff understand the basic methodologies underlying impact evaluation and identifying programs that could be rigorously evaluated. As an outcome of that meeting, many Foundation staff are currently exploring potential impact evaluations, including one that would look at the outcomes of alternative dispute resolution programs in Sri Lanka and another that would work with the Government of Malaysia to implement impact evaluations on their own health and social programs.

---

In addition, creating partnerships with experienced research organizations and conducting experimental evaluation are invaluable for implementing organizations. Independent organizations such as J-PAL, CEGA, and IPA have a deep pool of expertise that can be tapped for experimental design. In addition, these organizations can often tap into independent sources of funding for evaluations from research funders, so not all evaluations need to be donor-driven and financed. The Foundation has such a partnership with CEGA.

Lastly, and most importantly, implementing organizations must articulate a coherent evaluation policy and invest in the resources to back that policy. While decentralized efforts to enhance evaluation should be encouraged, they are often not enough to push evaluation as a priority, and by definition, will always be ad-hoc. Articulating a coherent policy helps to improve standards for evidence and set priorities for different types of evaluation within the organization. To ensure that such policies are created and institutionalized, implementers need to create separate evaluation units or committees with their own budgets. These internal units or committees would take on important tasks to help operationalize evaluation policy like establishing evaluation standards, linking projects with appropriate internal and external resources, helping design evaluations, and helping to organize and communicate results. The Foundation has done this through the creation of an Evaluation Task Force, which is charged with setting evaluation policy and helping to shape thinking about evaluation throughout the organization.

While such efforts would require considerable investment by implementing agencies, they are essential to build effective evaluation operations within an organization. For the Foundation and similar NGOs, it is important to begin learning more about program effectiveness.

4.2 Collective Action on Impact Evaluation

Moreover, collective action is needed across countries and agencies to push rigorous evaluation forward. As CGD (2006) notes:

“While specific actions by individual countries and agencies can reduce the evaluation gap, genuine progress is likely to be faster and more lasting if those countries and agencies collectively commit to increase the number and quality of impact evaluations.”

First and foremost, collective action could help donors and implementing partners make credible commitments to conduct rigorous evaluation outside of existing aid bureaucracies. This could help bypass some of the institutional disincentives that are present within such bureaucracies. In addition, donors and implementing partners could take advantage of economies of scale, while bearing fewer costs.

One way for donors to effectively do this is to commit to a common platform, in which each bureaucracy agrees to a set of common tasks. Such a commitment could piggyback
on already existing organization structures such as the High Level Forum on Aid Effectiveness, which will hold its fourth meeting in Busan, Korea in 2011. The 3rd High Level Forum held in Accra in 2008 committed to the principles of predictability, country ownership, use of country systems, conditionality based on country needs, untying aid, avoiding aid fragmentation, country partnerships and transparency. The principle of evaluation could also be added to this list. However, to work effectively, the goals, guidelines, and targets would have to be specifically spelled out so that there is less temptation for donors to shirk their responsibilities.

Another mechanism for collective action would be to build an independent committee or organization for evaluation to which donors are committed and contribute. As CGD (2006) points out, such an organization would also have the added benefit of having a wider membership than just donor agencies. This could include developing country governments, new or non-traditional donors, NGOs, the private sector and academia, and would contribute to an evaluation policy that takes into account the needs of all of these stakeholders. Core functions could include establishing evaluation quality standards, reviewing evaluation designs, providing grants for evaluations, “interpreting” and disseminating information, and maintaining databases.

In either case, a new mechanism for collective action could serve as a focal point for more rigorous impact evaluation, not only for donors, but for implementing partners as well. If donors credibly enter into such arrangements, it can serve as a strong signal to NGOs that donors take rigorous evaluation seriously and are willing to commit resources to it. And, of course, such commitments should eventually trickle down to implementing partners, who will be required to carry out such evaluations in some cases.

5. CONCLUSIONS

Even with a growing interest and commitment to rigorous impact evaluation, there are currently institutional and programmatic disincentives to rigorous impact evaluation for implementing partners. These disincentives are systematic and not due to any one cause, but a confluence of factors that permeate important stakeholders including donors and implementing partners.

However, these disincentives are not insurmountable. With smart investments in resources, policy, and training among donors and implementing partners, as well as collective action, the aid community can begin to use impact evaluation more effectively across a wider array of programs. This process has already been started among both donors and implementing partners, but greater efforts and investments must be made.

For NGOs like The Asia Foundation, more effective use of impact evaluations would be a boon. Impact evaluation can test the assumptions embedded in programs and approaches and have the potential to show where those assumptions are faulty. As one of the premier development organizations operating in Asia, program effectiveness is a chief concern, not only for programmatic feedback, but also in terms of what works best for its program partners and beneficiaries.
REFERENCES


14
www.3ieimpact.org/strategy/pdfs/3ie%20impact%20evaluation%20practice.pdf


