



*practical learning for  
effective international action*

***Working Papers on Program Review & Evaluation: #2***

# **Evaluating Impact in Conflict Prevention and Peacebuilding Programs**

**Mark M. Rogers**

***Reflecting on Peace Practice Program***

***2012***



## CONTENTS

---

About CDA Collaborative Learning Projects	i
About the Working Papers on Program Review and Evaluation	ii
I. The Notion of Impact	1
II. Impact Evaluability Assessments	2
III. Process Dimensions	3
IV. Standards in Evaluating Impact	4
V. Evaluation Questions and Analysis	5
VI. Conclusions	13
VII. The Growing Edge	13
References	15

## ABOUT CDA COLLABORATIVE LEARNING PROJECTS

---

### OUR VISION

CDA strives for a world in which people are supported to enhance their resilience, drive their own development and resolve conflicts without resorting to armed violence.

### OUR MISSION

CDA facilitates collaborative learning processes among humanitarian, peace, development and corporate practitioners and policy-makers to ensure that their support is effective and accountable.

### OUR PROGRAMS

The oldest of CDA's programs, **DO NO HARM (DNH)**, seeks to identify the ways in which international humanitarian and/or development assistance given in conflict settings may be provided in ways that avoid exacerbating the conflict and maximize potential contributions to peace processes. DNH's conflict sensitivity tools are now used by practitioners around the world to design and implement assistance efforts that improve the outcomes for affected communities.

The Reflecting on **PEACE PRACTICE PROGRAM (RPP)** is an experience-based learning process involving agencies whose programs attempt to prevent or mitigate violent conflict directly. One of the central insights from the RPP case studies is that, while individual peacebuilding efforts may be beneficial, they do not necessarily contribute to – or, in the aggregate, “add up” to – what RPP calls Peace Writ Large. The RPP aims to improve the effectiveness of peace efforts and their positive role in building durable peace.

The **LISTENING PROGRAM (LP)** is a comprehensive and systematic exploration of the perspectives and ideas of people who live in societies that have been on the recipient side of international assistance, including humanitarian aid, peacebuilding activities, or human rights work. Listening and responding to local perspectives is critical for donors and aid agencies to live up to the global commitments they have made to listen to and involve local stakeholders and to improve the quality, effectiveness and accountability of their efforts.

The **CORPORATE ENGAGEMENT PROGRAM (CEP)** collaborates with companies to ensure that they have positive rather than negative impacts on the communities where they operate. As part of this approach, the CEP assists companies in the development and implementation of practical options to build constructive relationships with the communities where they work. Ultimately, CEP supports the contribution of the private sector to stability and sustainable development.

## ABOUT THE WORKING PAPERS ON PROGRAM REVIEW AND EVALUATION

---

This Working Paper series is based upon findings regarding program review and evaluation processes from across all of the programs of CDA Collaborative Learning Projects (CDA). These include the Do No Harm, Reflecting on Peace Practice, Listening, and Corporate Engagement Programs.

All of CDA's programs are concerned with issues of impact, results, assessment and learning—as well as mechanisms for obtaining feedback and being accountable for program effects. This CDA Working Paper series aims to share reflections and lessons gleaned from across these programs for agencies who are undertaking their own program reviews and evaluations.

CDA produced these papers to offer practical support for practitioners seeking to evaluate the effectiveness and contributions of specific programs across several sectors. The papers do not represent final products and are working documents. CDA would like to acknowledge the generosity of the individuals and agencies involved in donating their time and experience and for their willingness to share their experiences and insights to generate and reflect on the learnings shared in these papers.

All CDA publications may be used, copied and distributed free of charge with appropriate acknowledgement and citation. In order to support our own ongoing learning and impact assessment processes, CDA welcomes your feedback and requests that you let us know how you are using our materials. Please e-mail your comments or feedback to [info@cdacollaborative.org](mailto:info@cdacollaborative.org).

### About the Author

Over the past six years Mark Rogers has helped peacebuilders use program evaluation to improve their practice. He has conducted external program evaluations in Ethiopia, DRC and the South Caucasus and internal evaluations in Burundi, DRC and El Salvador. He recently completed a one-year developmental evaluation of an early warning/early response collaboration in Liberia. He worked on context assessments in Chad, Niger, Mauritania and Kyrgyzstan. Mark co-authored a manual for peacebuilding practitioners, *Designing for Results: Integrating Monitoring and Evaluation in Conflict Transformation Programs*. He has taught courses on peacebuilding program evaluation at EMU, AU and for various NGOs in Africa and Asia. Mark has served in leadership positions with Search for Common Ground, Catholic Relief Services and recently the Life & Peace Institute.

## Evaluating Impact in Conflict Prevention and Peacebuilding Programs

This paper builds on the guidance developed by *OECD DAC* supporting the effective evaluation of peacebuilding and conflict prevention programs,<sup>1</sup> inputs from RPP consultations, an examination of actual evaluation practice and other sources. Its purpose is to promote deeper more meaningful discussions between peacebuilding evaluands, evaluators, donors and other evaluation users about impact evaluations and the distinctions between the effectiveness and impact criteria. It is not intended to be definitive or final, but rather an opportunity to deepen the discussion and practice of evaluation of peacebuilding work. The five core DAC evaluation criteria include relevance, effectiveness, efficiency, sustainability and impact. This paper focuses on impact evaluation.

Evaluation criteria and standards evolve through theory, practice, reflection, discussion, research, and meta-evaluation. This paper strives to identify different evaluation practices that can be employed to achieve more credible and useful evaluation findings relating to impact. This paper is not about preparing for an evaluation, developing terms of reference, selecting evaluators, etc. There are already numerous guides on these critical processes. The focus here is on the broad approach employed in impact evaluation. Although there are a number of established and more detailed references and guides, none of them are centered on peacebuilding. This reflection includes a first attempt at proposed standards against which peacebuilding evaluation designs and evaluation performance can be compared. We hope that it will result in some lively discussion and suggestions of alternative standards.

### I. The Notion of Impact

Impact evaluation assesses the changes in the conflict or context that can be attributed to an intervention, program or policy. Once we have established that the program was well implemented, had very positive effects, and achieved its intended outputs and outcomes, the remaining question is, “What changes in the conflict or the context are due to the program?”

Impact is relative and contextual. For some interventions, impact means macro level changes in the conflict or effects on *Peace Writ Large*. RPP suggests that impact should be at the peace writ large level – not necessarily the macro-level, but at least addressing the macro level factors and dynamics within the intervention’s sphere of influence. For other interventions, impact means changes at the level of the program goal, which may mean more local level changes in the conflict. This paper does not attempt to identify specific changes that constitute impact. These will vary from one context to another and from one program to another. For our purposes impact is understood to be the highest level change in the conflict or the context within the intervention’s sphere of influence. This creates some overlap and similarities with the effectiveness criterion. Impact evaluation does not just seek to

---

<sup>1</sup> References are based on language from *Evaluating Peacebuilding Activities in Settings of Conflict and Fragility—Improving Learning for Results* (Paris: OECD-DAC, November 2012). Provisional guidance on the evaluation of peacebuilding programs and activities was previously published in 2008 under the title, *Guidance on Evaluating Conflict Prevention and Peacebuilding Activities (Working draft for the application period)*.

identify changes in the conflict but to directly address the questions of causality and attribution.

We evaluate the impact of programs to learn what works in what contexts. This helps improve ongoing programs and can inform policy. Determining causality, or attribution, is at the center of impact evaluation. Most leading evaluation theorists recognize the need to combine qualitative and quantitative methods in assessing impact. Mixed methods, based on needs and constraints of the situation, are more likely to produce insightful and credible information than any single approach to impact evaluation.

## II. Impact Evaluability Assessments

Not all programs merit impact evaluation. Impact evaluation is most convincing where there is clear program theory that spells out why and how key changes actually come about. Programs that pair well-reasoned program design with results well supported by evidence are good candidates for impact evaluation. Impact evaluation establishes whether or not the changes that took place are attributable to the intervention. Theory-based evaluation explores how the intervention influenced those changes (Weiss, 1997, 2000, 2007; Davidson 2010). It makes little sense to invest in impact evaluation for programs where the central program theories are weak, inseparable from other causal factors, or too small to be of much significance.

Nor is impact evaluation feasible in every situation. Impact evaluations using randomized control trials, for example, are not very effective where the program cannot be manipulated in order to control for specific variables. Impact evaluations using more naturalistic methods may not be able to triangulate their findings in a conflict context. Regardless of the methods used, impact evaluation requires time, resources and considerable technical competence. It is not a add-on question to a different evaluation process that can be answered in a few hours or simply a matter of an evaluator's opinion.

The following list of questions, common in evaluability assessments, illustrates some of the issues to think about in considering whether or not to engage in impact evaluation.<sup>2</sup> The answers to these questions depend in part on when in the process they are asked.

### *Program Scope*

- Were the changes sought significant enough to influence *peace writ large* or some other macro-issue?
- Were the initial investments in peacebuilding sufficiently strategic and robust? Should we expect to see some ongoing effects over time?
- Were the processes or structures strengthened by the intervention(s) able to handle new and emergent conflicts? How did the initiative adapt to changing circumstances?
- How durable were the institutionalization components?

### *Contextual Considerations*

---

<sup>2</sup> For more information on evaluability assessment, please refer to the third paper in this CDA paper series, Working Papers on Program Review & Evaluation, titled "Evaluability Assessments in Peacebuilding Programming," by Cordula Reimann.

- For cyclical conflicts, have there been subsequent episodes of the same conflict that present opportunities for comparisons?
- Do the original strategies still fit the subsequent episodes and the changes in the context?
- Are the issues and populations the same (or comparable) across the different episodes?

#### *Timing*

- Will the intended evaluation be initiated early enough to allow for observation at different points in time? Has enough time elapsed to be able to see outcomes and their results?
- Has there been/will there be periodic evaluation (developmental, formative and summative) upon which to build?
- Will the timing be recent enough to facilitate recall and cover enough time to matter?

#### *Value for Money*

- Will the answers to the evaluation questions justify the cost of the evaluation?
- Does the programming in question represent an important direction within the larger organization or the peacebuilding work in the location selected?

Where evaluability assessments reveal weak cases for impact evaluation, different evaluation questions relating to relevance, outcomes and implementation rather than impact may be in order. The focus of this paper is on evaluation processes rather than specific impacts of interest to the peacebuilding community. No effort is made here to prescribe universal indicators of peacebuilding impact.

### **III. Process Dimensions**

A comprehensive design for impact evaluation sets out how the evaluation will perform the full range of evaluation tasks. Building on what the Network of Networks of Impact Evaluation (NONIE) Subgroup 2 maintain (NONIE 2008, 6), those tasks are:

- Comprehensive identification of important impacts - changes in the context and/or conflict; (negative, positive, intended and unintended and their durability over time)
- Systematic and defensible data collection and analysis of evidence of these impacts;
- Sound inferences about the contribution of the intervention to achieving these impacts; and
- Effective management of the evaluation, including transparent reporting of methodology and, where appropriate, formal meta-evaluation.

A subsequent NONIE guidance in 2009 builds on the earlier work and outlines nine key issues in impact evaluation. The NONIE issues imply steps or a sequence of evaluation activities that are worth keeping in mind when designing impact evaluations. The 2009 NONIE issues (Leeuw 2009, xxi) are as follows:

1. Identify the type and scope of the intervention.
2. Agree on what is valued [by different stakeholders].
3. Carefully articulate the theories linking interventions to outcomes.
4. Address the attribution problem.
5. Use a mixed-methods approach: the logic of the comparative advantages of methods.
6. Build on existing knowledge relevant to the impact of interventions.



7. Determine if an impact evaluation is feasible and worth the cost.
8. Start collecting the data early.
9. Front-end planning is important.

## IV. Standards in Evaluating Impact

*“First to be more effective, the standards require adaptive, responsive and mindful use. They should not be applied literally and superficially following a simple recipe. Rather, evaluators must discover how to apply them adaptively in each specific situation to achieve evaluation quality” (Yarbrough 2011).*

The draft standards proposed here are offered in the same spirit. Standards are intended to assist in the pursuit of evaluation quality, rather than to provide exact procedures for rigid application in every situation. Standards can be adjusted upwards for advanced and sophisticated initiatives. However, some minimal thresholds are helpful in establishing credibility.

The dimensions in the following table come from a 2008 NONIE initiative and a paper by Patricia Rogers and a large number of other contributors. Most of the standards were included in the subgroup’s paper and a few others have been added here.

<b><i>Dimensions</i></b>	<b><i>Proposed Standards</i></b>
<i>I-1 Comprehensive identification of important impacts</i>	a) Open-ended, goal-free exploration of changes and developments in the conflict and the context b) Look for both negative and positive impacts c) Look for both intended and unintended impacts d) Durability of changes over time?
<i>I-2 Systematic and defensible data collection and analysis</i>	a) Situationally appropriate methods b) Appropriate combination of evaluation methods c) Triangulation (data, theory and methodological)
<i>I-3 Sound inference of contributions of the interventions</i>	a) Other variables controlled or methodically addressed b) Comparisons made where possible c) Consideration of program theory d) Inference based on evidence e) Conclusions stem from a generally recognized process(es) of inference
<i>I-4 Effective evaluation management</i>	a) Fit with existing knowledge b) Transparency in reporting c) Methodical review of threats to validity, reliability & objectivity d) Use of meta-evaluation

Each of the dimensions and the associated standards listed in the table above are explored below.

## V. Evaluation Questions and Analysis

### *I-1 Comprehensive Identification of Important Impacts*

#### **Rationales and Clarification**

Programs and interventions sometimes lead to unintended or to negative results. Evaluations that only look at the program or its participants or beneficiaries or look only at intended and positive changes are unlikely to detect other less positive changes. One way to extend beyond the boundaries of the program is to begin with open-ended or goal-free inquiry focused on changes in the conflict and/or context. This can be done by completing a conflict analysis or updating an existing conflict analysis. The intent is to identify changes in the conflict and the context over the period of time covered by the evaluation, without specific consideration of the intervention. Because the evaluation is not looking at the changes through the lens of the program, it is more likely to learn of negative and unintended consequences as well as positive and intended consequences.

The first step involves inventorying all the impact-level changes in the conflict and the context, as well as those specified by the program or the intervention. These are then classified according to intent and whether or not they are positive, negative, or ambivalent vis-à-vis the conflict. Some assessment is also made regarding the anticipated durability of the impact over time, which is closely related to the OECD DAC sustainability criterion.

Many peacebuilding evaluations do not progress beyond identifying the changes that the program may have influenced and fail to establish either contribution or attribution. While helpful, simply identifying the important impact level changes that took place does not qualify the evaluation as an impact evaluation.

#### **Applying Comprehensive Identification of Impacts**

An in-depth analysis of a cyclical conflict in X country revealed that street violence by university students preceded all five major outbreaks in violence over the eleven years prior to the program. An NGO set up a program to help youth, particularly political youth, to deal effectively with efforts to manipulate them into violence. Political youth included university-based student unions, youth cadres within the established political parties and youth groups within governance processes, such as the youth advisory board to the city council.

The program activities included a five-day experiential outdoor challenge, in which groups consisting of people from diverse backgrounds build their collaboration and team skills in order to overcome challenges in the natural world, followed within several months by a five-day course in governance, representation, leadership, conflict analysis and conflict transformation. These two components of the program reached over 750 youth in three provinces. One hundred twenty graduates from these two initiatives, who are involved with groups that reach across traditional ethnic divides, enrolled in a mentorship program with a conflict transformation coach. A national hot-line was set up to provide program participants with access to other participants from different groups and access to technical assistance with conflict transformation. The hot line averaged 35 calls per month, of which 7 per year resulted in professional technical assistance by program staff over the last three years.

Four years after the program began and one year before it was to end, the NGO and its donors wanted to know:

- What worked in reducing the effectiveness of political manipulation of youth? For whom did it work and with what results?
- Have the relationships developed over the course of the program translated into ongoing relationships active in peacebuilding?

The evaluation team began by updating a recent conflict analysis done by a well-regarded peace research agency, both as a means to understand the relevance of the program and to identify changes in the conflict of the context.

In addition, the evaluation team used the Most Significant Change (MSC) process to identify changes in the context and in the conflict. MSC participants went through several sorting and prioritization rounds, but did not select a single most significant change. Evaluators duly recorded the criteria participants used in making their choices as well as the actual stories.

MSC was accompanied by interviews with local authorities, informed observers, non-political youth and political youth and other key stakeholders to identify changes in the conflict and context since the program started. Each interview began with open-ended questions and then focused on both positive and negative questions and ended with intended and unintended results.

The primary changes found were:

- On three occasions, political operatives unsuccessfully attempted to mobilize youth groups to violence and, on another occasion, initial youth violence triggered by political agitators was quickly contained by different youth groups working in collaboration.
- Almost all political youth who participated in the program had been marginalized within their own parties. They reported a loss of influence stemming from their raising sensitive and unpopular questions and their associations with youth from other parties.
- Five out of 350 youth engaged in youth advisory boards that participated in the program ran for municipal office in the year following the end of the program. Three were elected – one from the progressive party and two from two different conservative parties.

## ***1-2 Systemic and Defensible Data Collection and Analysis***

### **Rationales and Clarification**

The evaluation needs to be open enough to detect unintended consequences and negative results. It needs to go beyond a narrow test of a specific, intended positive result. Defensible data collection and analysis generally includes some combination of quantitative and qualitative methods. Mixed methods “strengthen the quality of evidence by compensating for the limitations of any one source through a complementary mix of research questions, processes for developing research questions, sampling procedures, data collection procedures, type of data and/or type of data analysis” (NONIE 2008). The evaluation should detail the rationale for the data collection and analysis methods used.

Establishing causality may require relatively sophisticated data collection methods. For example, the

International Initiative for Impact Evaluation (3ie) outlines much more specific standards for systemic and defensible data collection and analysis for its approach to impact evaluation:

“3ie supports impact evaluations that adhere to agreed-upon methodological standards for addressing the “attribution challenge” – e.g. establishing cause and effect between programmatic activities and specified outcomes. In particular, evaluation designs must be capable of addressing: a) confounding factors; b) selection bias; c) spillover effects; d) contamination of control groups; and e) impact heterogeneity by intervention, beneficiary type and context.”<sup>3</sup>

Explaining these methods exceeds the scope of this paper. They are mentioned in order to alert the reader to some of the complexities typically found in impact evaluation.

The rationale for choosing certain data collection and analysis methods over others should consider the situation or context in which the program and the evaluation take place. Often, peacebuilding programs encounter contextual constraints to gathering and analyzing data, such as sensitivity of the issues, access to difficult to reach stakeholders, vested interests, political interference, etc. Other constraints in evaluation include time, budget and utility constraints.<sup>4</sup> Often these are set out in the terms of reference without considering other ways of gathering data and doing data analysis. Evaluators may need to use acceptable but less optimal methods in order to work within these types of constraints.

The tools and techniques for data collection and analysis are inherently linked to those for making sound inference. These are listed in section I-3 below.

### **Applying Systemic & Defensible Data Collection and Analysis**

How should the evaluation gather and analyze data that would demonstrate the contribution of the program toward the four occasions where political manipulation of youth had been blocked or contained? There were several challenges in considering an impact evaluation design that would be appropriate in this context:

- Youth were mobile –they frequently changed their e-mail addresses and were difficult to track one year or more after their involvement with the program.
- Resilience against manipulation is often complex and involves multiple interdependent variables.
- There were numerous similar and different peacebuilding initiatives in the same places, over the same time period, and most participants had engaged in or heard of at least one other program.

The program area included all conflict-prone communities in the country. There were no similar communities to serve as control groups. Non-similar control groups were rejected because the

<sup>3</sup> For more information on 3ie’s Principles for Impact Evaluation, please refer to [http://www.3ieimpact.org/media/filer/2012/05/17/principles\\_for\\_impact\\_evaluation.pdf](http://www.3ieimpact.org/media/filer/2012/05/17/principles_for_impact_evaluation.pdf), last accessed November 15, 2012.

<sup>4</sup> For options within budget constraints see Bamberger *et al.* 2004.

evaluation could not control for a variety of pre-existing differences. These circumstances and budget limitations required an alternative to experimental approaches.

The evaluation team employed a pre-test/ex-post-test comparison of the 120 participants in the mentoring program. This was anticipated during the formative evaluation, and those results served as a base against which to compare ex-post responses. During the final year of the program efforts to locate the original 120 participants resulted in responses from 57 participants.

The evaluation also developed detailed case studies regarding each of the four prevention stories. A multi-disciplinary team including a sociologist, a psychologist and a political scientist prepared the case study guidance. Snowball sampling led to different stakeholders and some of the lower level protagonists in these events, but it quickly became impossible to access the principal instigators. Nonetheless, the case studies provided details on how youth mobilized, which youth mobilized, what motivated them to act and the sources of their ideas and strategies.

The evaluation also drew upon the data in the log of hot line calls made over the course of the program; disaggregated by school year, date of the call and by the number of years the callers had studied at the university. They compared the hot-line log with a time-line showing when manipulation was most intense.

### ***1-3 Sound Inference Regarding the Intervention's Contributions***

#### **Rationales and Clarification**

Inference is used to explain causal links between an intervention or treatment and its effects. The scale and scope of the interventions pose different challenges for inferring causation.

As the NONIE authors explain, “inferring causation clearly becomes increasingly complex across continuums that extend from:

- Outputs to outcomes to impact, with progressively weakening reliability of the programme theory and hence the certainty of attribution;
- Projects (usually characterized by single interventions with explicit objectives and change effects measured in the short term) to comprehensive programmes where activities cut across sectors, themes or geographic areas, with a number of confounding variables, and change effects measured in the longer term; and
- ‘Simple, single strand interventions’ to ‘complex adaptive systems’ characterised by large numbers of unknown variables and unknown causal connections between variables, by interactions, feedbacks and nonlinear relationships, and high sensitivity to small perturbations” (NONIE 2008).

To deal with the different ways in which interventions contribute to impact, evaluators must rely on both quantitative and qualitative methods. The methods and the mix of methods need to be thoughtfully explored and negotiated with the evaluation team at the point when the evaluation is being designed.

The use of counterfactuals, which is central to the experimental approach, compares control or

comparison groups who did not receive the treatment with those that did, while controlling other variables. The use of counterfactuals is one way to address the question, “What would have happened in the absence of the intervention of interest?” Or, where comparison groups exist, the aim is to address the question “What did happen in the absence of the intervention of interest?”

In addition to the classic experimental design, there are a number of quasi-experimental techniques that offer less convincing but still valid claims, which may still be useful to policy-makers and program managers. These generally attempt to either approximate randomness and/or use comparison and non-equivalent comparison groups as opposed to control groups. Quasi-experimental methods mimic randomization by using propensity score matching or multivariate regression and mimic comparison groups using the difference-in-difference approach.

There are also qualitative methods of inferring causal relationships between variables. Patton (2010) offers empirical data with source triangulation as a viable alternative. Scriven offers *modus operandi* and *causal list* methods to argue for or against attribution. Miles and Huberman (1994) provide a more nuanced understanding and offer an expanded taxonomy of nine different types of causal relationships between variables.<sup>5</sup> Qualitative methods rely on multiple lines and levels of evidence. Specific techniques include participatory performance story telling and process mapping, to name a few.

Regardless of whether an evaluation is using experimental designs, quasi-experimental techniques of naturalistic methods, or hopefully some combination of the two, the evaluation should include the evidence upon which inferences are made and what additional information, if any, would help support the inferences made.

### **Key Questions Relating to Sound Inference**

- Were other possible causal explanations or scenarios explored and convincingly eliminated?
- Was the program theory strong enough to merit impact evaluation?
- What was likely to have happened (or did actually happen) in the absence of the intervention?
- How did the evidence support the inferences made?
- How did the inferences draw on established practices/methods for making inferences?
- Were the threats to validity recognized and addressed?

---

<sup>5</sup> Examples include: temporality, strength of association, biological gradient, coherence and analogy, among others.

## Applying Sound Inference

Each case study traced the prevention process and the roles played by youth (program participants and non-participants). There were some commonalities. Notably, in all four events, youth leaders from various groups immediately reached across divisions to engage other youth groups, deny any alleged support for violence and express their solidarity with each other. The most recent case took place on a university campus and did not involve any of the youth who had gone through the program, since the last participants had graduated from the university in the final year of the program. Further examination revealed that both the number of program participants and their level of influence decreased over time with each incident.

When asked about their motivation to intervene, over half of the youth (both participants and non-participants) shared some version of a story of a small group of young peacebuilders who, ten years earlier, had effectively intervened in an escalating conflict on campus, during one of the country's most volatile moments. The story was widely known prior to the program, and the program had further disseminated this story through its classroom curricula, its outreach efforts and radio spots. When asked where they first learned of the story, over half made reference to some dimension of the program. Nowhere in the many incarnations of the story heard during the evaluation was any mention made of immediately reaching across divisions. Apparently the story served more as a motivator than a road map. So where did the strategy of immediately reaching across divides come from?

Further inquiry revealed that the of 85 young activists interviewed for the cases studies, 37 believed that the strategy of reaching across divide was spontaneously developed in the moment, 20 made references to materials they learned in the program, 15 made reference to materials not covered in the program, and 13 did not know.

The evaluation identified and explored additional causal explanations for the four prevention outcomes of interest. The following three potential causal explanations could not be discounted.

1. The program coincided with the introduction of the Young Peacebuilder position in the indigenous dispute resolution circles. This carried considerable prestige and social status. Most of the participants were university students and graduates. These leaders were spread throughout different youth groups and exercised considerable influence, especially during the first three unsuccessful attempts at manipulation. Thirty-five of the program participants were also Young Peacebuilders. The Young Peacebuilder curriculum and the curriculum of the evaluand promote early communications across divides as a means to contain the situation and avoid further escalation.
2. Most political observers maintained that the first and third efforts to manipulate youth into violence were organized by low-level party members who lacked the clout and mobilization skills needed to be effective.
3. In the second case, an inter-agency early warning system had anticipated the event and informed university youth who were on alert far enough in advance for them to contain the events. These alerts coincided in time with increased use of the program hot-line. Hot-line calls doubled during the days immediately prior to and during the four occasions where manipulation was blocked or contained.

Given that these additional causal explanations that could not be discounted, the program was not a good candidate for impact evaluation.

## ***I-4 Effective Evaluation Management***

### **Rationales and Clarification**

Given that the research base in peacebuilding is still nascent, many evaluations overlook an important preliminary question. Ravillion asks, “What do we currently know about this type of intervention and what are the most important knowledge gaps?” (in Chambers 2009, 15) Understanding how evidence has been used to establish causality in the past is a helpful starting point.

The evaluation methods need to be transparent. They should be clearly detailed in the evaluation reports.

Managing the evaluation process should include periodic and methodical review of any threats to validity, reliability and credibility. Overall evaluation validity extends beyond the issue of data collection to include perceptions about the evaluation, the evaluators, the processes used, the design and the ways findings are developed, shared and used. It includes the trustworthiness of the evaluators and their credibility in all the major evaluation activities. Key quality control and assurance activities focus on managing human resources, preserving the integrity of data, managing external communications and relations, and protecting information and safeguarding identities that may be at risk.

Evaluations, like programs, benefit from evaluation. Meta-evaluation involves the evaluation of an evaluation – any type of evaluation, not just impact evaluations. This can happen at various times in the evaluation process. For example, it can be helpful to assess the strengths and weakness of evaluation design, prior to any data collection. By assessing design weaknesses prior to data collection, the evaluators can strengthen the design, increase utility and conserve resources. A critical friend, an evaluator who is not part of the evaluation, can be very helpful in assessing evaluation design issues as well as actual evaluation processes and methods applied during the evaluation. Meta-evaluation, like most evaluations, has the most to offer when it is initiated while the evaluation is still in progress.

How do we determine if an evaluation is well done? Or as Scriven (2011) asks, “What are the criteria of merit for an evaluation?” Scriven offers a menu of options in conducting meta-evaluations:

#### *Validity*

- Meta evaluation parameters: All evaluations need parameters involving contextual considerations, the focus and objectives of the meta-evaluation, levels of analysis, level of detail, etc.
- Accuracy: The logic soundness and the adequacy of evidence, including evidence of confirmation
- Reliability: Consistency between several sources
- Comprehensive coverage of all the relevant considerations

*Clarity:* Stakeholders and audiences can comprehend the report, process and findings

*Credibility:* Low levels of apparent bias



*Propriety: Ethically and culturally appropriate*

*Cost Considerations/Value for Money*

### **Applying Effective Evaluation Management**

The head of the M&E department of the NGO headquarters asked the country program, the evaluation team and the regional M&E advisor to participate in a meta-evaluation focused on the evaluation process and more particularly on reliability, comprehensive coverage of all relevant considerations and credibility. Two impact evaluations of similar programs were planned for the following year and would benefit from the learning about impact evaluation from this program. Together they developed the terms of reference for the meta-evaluation and recruited an evaluator.

The meta-evaluator began work during the last year of the program, only slightly later than the evaluation team, but while the evaluation team was still working. She was able to attend several of the meetings on the case studies, as well as observe a few of the evaluation data analysis sessions. She interviewed members of the evaluation team, program participants and evaluation users. She also facilitated a reflection workshop with the evaluators and the county program staff. The meta-evaluator was not asked to comment on the content of the findings, only on the legitimacy of the evaluation practices and processes.

Findings, discussion points, and recommendations were captured in a meta-evaluation report.<sup>6</sup>

Based on the findings of the meta-evaluation, the initial meta-evaluation group developed the evaluation guidance for the upcoming impact evaluations.

## **VI. Conclusions**

Impact ranks among one of the most commonly requested evaluation criteria among the OECD DAC criteria. It is also among the least applied, due to the confusion between impact and effectiveness criteria. Attribution and causal inference appeal to program staff, as well as fundraisers and donors. There are a multitude of quantitative and qualitative tools that can be employed in efforts to establish causal relationships between program interventions and changes in the conflict and/or context. Choice of methods needs to be situationally appropriate and viable within data, time, budget and contextual constraints and technical competence. Alternative explanations need to be considered and either eliminated or controlled for, where possible. Inference needs to consider existing knowledge of the variables in question, be evidence-based, and at least reference if not conform with established logic in making inferences. Methods will vary according to the complexity of the conflict and scope and complexity the interventions.

Peacebuilding evaluations charged with evaluating impact routinely raise the inherent challenges in demonstrating attribution and often find the challenges insurmountable. There is an emerging consensus within peacebuilding that contribution analysis, may be as good as it gets until considerable

---

<sup>6</sup> For an example of one approach to meta-evaluation see Action Asia Practitioners' Forum (Action Asia 2011).

more peacebuilding research is available.

“Contribution analysis examines a postulated theory of change against logic and evidence to test the theory of change including examining other potentially influencing factors that could explain the observed results. The overall aim is to reduce the uncertainty about the contribution the program is making to the observed result. The result of a contribution analysis is not definitive proof that the program has made an important contribution, but rather evidence and argumentation from which it is reasonable to draw conclusion about the degree and important of the contribution, within some level of confidence. The aim is to get a preponderance of evidence, as in the judicial tradition” (Patton 2008).

Patton contends that contribution analysis is particularly appropriate in situations where multiple initiatives are working toward the same end, where impacts occur over long periods of time, and where “simple notions of linear attribution are neither meaningful nor accurate” (Patton 2008).

The field of peacebuilding is going to need to develop and demonstrate greater competence in both impact evaluation and contribution analysis in order to better understand the connections between interventions and effects.

## VII. The Growing Edge

In this effort to contribute to a deeper and meaningful discussion about impact in peacebuilding evaluation, several questions remain unanswered.

- In what peacebuilding situations are the different impact evaluation methods most appropriate?
- What are important precedents of contribution analysis in peacebuilding evaluation?
- How does conflict sensitivity figure into impact evaluation?
- What are some of the non-linear frameworks for mapping common peacebuilding contributions in complex and dynamical situations?
- How can we encourage meta-evaluation (an evaluation of the evaluation) and the sharing of meta-evaluations of impact evaluations of peacebuilding interventions that tend to be internal documents?
- Are there general categories of changes/types of changes that could be indicative that an initiative has had an impact on peace writ large in different types of contexts?

## REFERENCES

- Action Asia Practitioners' Forum. 2011. *Examining RPP as a Tool for Evaluation*. Colombo: Action Asia.
- Bamberger, Michael, Mary Church, Lucia Fort, and Jim Rugh. 2004. "Shoestring Evaluation: Designing Impact Evaluations under Budget, Time and Data Constraints." *American Journal of Evaluation* 25(1): 5–37.
- Chambers, Robert, *et. al.* 2009. "Designing Impact Evaluations: Different Perspectives." International Initiative for Impact Evaluation (3ie) Working Paper #4. New Delhi: 3ie.
- Dart, Jessica and Rick Davies. 2005. *The 'Most Significant Change' (MSC) Technique: A Guide to Its Use*. Melbourne: MandE. <http://www.mande.co.uk/docs/MSCGuide.pdf>, last accessed November 15, 2012.
- Davidson, Jane E. 2010. "Ascertaining Causality in Theory-Based Evaluation," *New Directions in Evaluation*. 87:17-26
- Gertler, Paul J., *et. al.* 2010. *Impact Evaluation in Practice*. Washington, DC: World Bank. <http://web.worldbank.org/WBSITE/EXTERNAL/EXTABOUTUS/ORGANIZATION/EXTHDNETWORK/EXTHDOFFICE/0,,contentMDK:22796485~pagePK:64168445~piPK:64168309~theSitePK:5485727,00.html>, last accessed November 15, 2012.
- Hughes, Karl and Claire Hutchings. 2001. "Can we obtain the required rigor without the randomization? Oxfam GB's non-experimental Global Performance Framework." International Initiative for Impact Evaluation Working Paper #13. New Delhi: 3ie. <http://www.3ieimpact.org/en/evaluation/working-papers/working-paper-13/>, last accessed November 15, 2012.
- International Initiative for Impact Evaluation (3ie). Undated. "Principles for Impact Evaluation." Washington, DC: 3ie. Available at <http://www.3ieimpact.org/en/about/strategy-documents-reports/>, last accessed November 15, 2012.
- Leeuw, Frans and Jos Vaeseen. 2009. *Impact Evaluation and Development: NONIE Guidance on Impact Evaluation*. Washington DC: NONIE—The Network of Networks on Impact Evaluation.
- Miles, Mathew and Michael Huberman. 1994. *Qualitative Data Analysis, an Expanded Sourcebook*. Second Edition. Thousand Oaks: Sage Publications, Inc.
- NONIE Sub Group 2. 2008. "Impact Evaluation Guidance, section 1, NONIE Subgroup 2."
- OECD Development Assistance Committee. 2012. *Evaluating Peacebuilding Activities in Settings of Conflict and Fragility—Improving Learning for Results*. Paris: OECD DAC. Available at

[http://www.oecd-ilibrary.org/development/evaluating-donor-engagement-in-situations-of-conflict-and-fragility\\_9789264106802-en](http://www.oecd-ilibrary.org/development/evaluating-donor-engagement-in-situations-of-conflict-and-fragility_9789264106802-en), last accessed November 15, 2012.

- 2008. *Guidance on Evaluating Conflict Prevention and Peacebuilding Activities: Working draft for application period*. Paris: OECD DAC Network on Conflict, Peace and Development Co-operation and the DAC Network on Development Evaluation. Available at <http://www.oecd.org/dac/evaluationofdevelopmentprogrammes/dcdndep/publicationsanddocuments.htm>, last accessed on November 15, 2012.

Patton, Michael Quinn. 2011. *Developmental Evaluation: Applying Complexity Concepts to Enhance Innovation and Use*. New York: The Guilford Press.

- 2008. *Utilization-Focused Evaluation: The New Century Text*. 4th edition. Thousand Oaks: Sage Publications.

Rogers, Patricia, and Carol Weiss. 2007. "Reflections ten years on: Theory-based evaluation: past, present and future." *New Directions in Evaluation*. Vol. 2007, Issue 114: 63-81.

Scriven, Michael. 2011. "The Meta-evaluation Checklist." <http://michaelscriven.info/images/EvaluatingEvals-Checklist.pdf>, last accessed November 15, 2012.

Stake, Robert E. 2010. *Qualitative Research: Studying How Things Work*. New York: The Guilford Press.

Weiss, Carol H. 2000. "Which Links in Which Programs Shall We Evaluate?" *New Directions in Evaluation*. Vol. 2000, Issue 87: 35-45.

- 1997. "Theory-based evaluation: past, present and future." *New Directions in Evaluation*. Vol. 1997, Issue 76: 41-55.

Yarbrough, Donald B. et. al. 2011. *The Program Evaluation Standards*. Thousand Oaks: Sage Publications.