UNITED STATES INSTITUTE OF PEACE www.usip.org

SPECIAL REPORT

2301 Constitution Ave., NW • Washington, DC 20037 • 202.457.1700 • fax 202.429.6063

SPECIAL REPORT 280   JUNE 2011

CONTENTS

Andrew Blum

Improving Peacebuilding Evaluation

A Whole-of-Field Approach

Summary

- The effective evaluation of peacebuilding programs is essential if the field is to learn what constitutes effective and ineffective practice and to hold organizations accountable for using good practice and avoiding bad practice.

- In the field of peacebuilding evaluation, good progress has been made on the intellectual front. There are now clear guidelines, frameworks, and tool kits to guide practitioners who wish to initiate an evaluation process within the peacebuilding field.

- Despite this, progress in improving peacebuilding evaluation itself has slowed over the past several years. The cause of this is a set of interlocking problems in the way the peacebuilding field is organized. These in turn create systemic problems that hinder effective evaluation and the utilization of evaluation results.

- The Peacebuilding Evaluation Project, organized by USIP and the Alliance for Peacebuilding, brought funders and implementers together to work on solutions to the systemic problems in peacebuilding work. This report discusses these solutions, which are grouped into three categories: building consensus, strengthening norms, and disrupting practice and creating alternatives. Several initiatives in each of these categories are already under way.

In May 2010, the Alliance for Peacebuilding launched the Peacebuilding Evaluation Project (PEP) in collaboration with the United States Institute of Peace. Over the course of a year, the project held a series of four meetings in Washington, DC.[1] The goal of the project was straightforward: to foster collaboration among funders, implementers, and policymakers in order to improve evaluation practice in the peacebuilding field. Participants came from a range of non-governmental organizations (NGOs), government funders, private foundations, the United Nations, and the U.S. government.[2]

This report is inspired by the deep and far-ranging conversations that took place at the four meetings of the project.[3] It reaches two central conclusions: First, improving evaluation must be addressed as a problem of structural and institutional change within the peacebuilding field, not simply as a technical research or methodological problem; and second, for this reason, whole-of-field approaches are necessary to make progress in peacebuilding evaluation. To support these conclusions, the report identifies a set of four fieldwide problems that hold back progress in peacebuilding evaluation and describes how these problems create self-perpetuating and problematic dynamics. The report then describes solutions participants identified for these vicious circles, as well as initiatives that meeting participants have already undertaken that contribute to whole-of-field solutions.

Definitions

Definitions of concepts such as peacebuilding or evaluation are always controversial, and should be, as the debates over these words are useful in pushing the thinking of the field forward. The purpose of this report, however, is not to engage in conceptual work, so it will rely on standard definitions in the field. Regarding peacebuilding, the Organization for Economic Cooperation and Development (OECD) provides a good working definition, which includes activities designed to prevent conflict through addressing structural and proximate causes of violence, promoting sustainable peace, delegitimizing violence as a dispute resolution strategy, building capacity within society to peacefully manage disputes, and reducing vulnerability to triggers that may spark violence.[4]

Regarding evaluation, a relatively broad consensus has emerged around a core set of documents that provide guidance on how to conceptualize peacebuilding evaluation. The 2008 OECD Fact Sheet on evaluating peacebuilding states that "evaluation assesses the merit and worth of an activity. . . . This learning process helps ascertain the quality of policies and programs, enhance performance, identify good practices, and define appropriate standards."[5] Cheyanne Church and Mark Rogers argue that "evaluation is the systematic acquisition and assessment of information gathered on specific questions to provide useful feedback for a program, organization, or individual. . . . Evaluation is commonly thought to serve two purposes: learning and accountability."[6] John Paul Lederach, Reina Neufeldt, and Hal Culbertson use the phrase *reflective practice* instead of evaluation and frame the concept as a suggestion to practitioners to reflect explicitly on "how things work," on what they have learned from experience, and on developing experience-based theory.[7]

While the above approaches are quite different in many ways, three key themes emerge from the documents. First, evaluation needs to be based on systematic collection of evidence. Second, evaluation serves both learning and accountability purposes. Third, evaluations are part, but not the entirety, of the evaluation process, which includes a range of monitoring and assessment activities that take place throughout a project's life cycle. Thus, as a starting point, we define peacebuilding evaluation here as an evidence-based process designed to create accountability for and learning from peacebuilding programs.

Slowing Progress at a Crucial Time

In 2008, Cheyanne Church argued that peacebuilding evaluation had moved from infancy to adolescence, and like any teenager, it "is full of untapped potential that needs to be supported and nurtured if it is to develop."[8] Church then suggested several milestones regarding the field's path toward maturity, including transparency, improved monitoring of programs, improved ability to compare across programs, and the professionalization of evaluation.[9] By these measures, unfortunately, the field has not matured significantly since 2008.

And yet for many reasons, this is a crucial time for peacebuilding evaluation. Peacebuilding and peer fields such as international development and democratization are under increased pressure to demonstrate their effectiveness.[10] At the same time, expectations for what counts as evidence of effectiveness have risen. There is demand for increasingly rigorous evaluation and a pushback against success stories and other forms of unsubstantiated claims.[11] Noting these pressures, one PEP participant said, "We need to become a learning field. To base our decisions on evidence rather than gut feelings."[12]

So why has the maturation process slowed since 2008? It was this question that provided the impetus for the establishment of the Peacebuilding Evaluation Project. And it was the provisional answer that informed how the initiative was designed. The premise of the PEP and this report is that progress in the peacebuilding evaluation field has slowed because much of the low-hanging fruit is gone.

Imagine an individual working in the peacebuilding field in 2000 who wanted to improve peacebuilding evaluation. A researcher could put together an evaluation framework, develop training manuals, and do research on methodology. Practitioners could use—or, more likely, hire a consultant to use—one or more of these frameworks or manuals in individual programs, or organize trainings for staff on those frameworks. These kinds of activities, according to Church's metaphor, marked peacebuilding's growth into adolescence.

The low-hanging fruit, now picked, is much of the intellectual and technical work that has been done regarding how peacebuilding evaluation should be conducted, and some small-scale evaluations that have been conducted primarily at the project level. There is now a collection of intellectual work that, in the aggregate, contains good strategies for peacebuilding evaluation, relatively solid guidance on what strategies should be used, and advice on how various strategies should be implemented.[13] Approaches from the larger development field include the outcome mapping approach developed by the International Development Research Centre,[14] the most significant change approach,[15] and randomized controlled trials (RCT) of the kind that the Poverty Action Lab at the Massachusetts Institute of Technology promotes.[16] Evaluation processes have not necessarily become increasingly successful because of this intellectual and technical work; in many cases, they have yielded unsatisfying results, but this is not caused by a lack of a knowledge base regarding strategies for effective evaluation.

Because much of the low-hanging fruit has been picked, the marginal value of further intellectual and technical work is diminishing. More needs to be done, but the existing frameworks and manuals are good enough, so that radically improving on them in the short term will be difficult. Similarly, project-level evaluations and capacity-building exercises should continue to be conducted, but increasing the value of these initiatives depends on changes at a larger, systemic level. Such changes, however, have been hard, and based on the discussions within the PEP group, it appears that this is so because the peacebuilding field has a number of self-perpetuating dynamics that create dysfunctional practices regarding peacebuilding evaluation. These vicious circles resist change and hold back progress. The self-perpetuating dynamics also create the need for a whole-of-field approach. Individual organizations working on small-scale project evaluations or improved evaluation methodologies will continue to struggle to advance the field unless more systemic problems are addressed. The rest of this report articulates what these broader problems are and describes solutions that emerged from the PEP discussions.

Four Problems

This section describes four interrelated sets of problems that affect the practice of peacebuilding evaluation: the scale problem, the weak results problem, the accountability chain

Progress in the peacebuilding evaluation field has slowed because much of the low-hanging fruit is gone.

problem, and the request for proposals (RFP) problem. It is the interactions of these problems that create the vicious circles that are holding back progress in peacebuilding evaluation.[17]

It is important to note that these problems are not specific to the peacebuilding field; one accomplishment of the intellectual work in peacebuilding has been to create a more sophisticated understanding of what challenges are common to the evaluation of social programs in general and what challenges are unique to peacebuilding. This report uses the term *peacebuilding evaluation,* but the issues discussed below are not necessarily unique to peacebuilding work.

The Scale Problem

The scale problem emerges because most work in the peacebuilding field is done as part of projects. These projects often focus on a particular area or topic, and while they can last for up to five years, most only last one or two. However, what people consider meaningful peacebuilding results emerge at a scale that is larger than the project level. In the first PEP meeting participants identified this as the tension between the project level and "peace writ large."[18] Scale tensions can emerge, however, even below the level of peace writ large, as peacebuilding efforts even in a single community are one piece of a much larger puzzle. Imagine an organization running a series of five intergroup workshops in a community that experiences a partial shift toward peace. Even if the workshops are successful, the complexity of any shift toward peace means that the workshops are only part of a set of much broader processes.[19]

The key challenge here is not how peacebuilding activities can achieve the goal of peace writ large; CDA Collaborative Learning Projects is doing important work on this with their case studies on cumulative impact.[20] Instead, the problem involves how the mismatch between project-level results and outcomes that are broad enough to be termed peace, by any consensus definition, affect the practice of peacebuilding evaluation.

The Weak Results Problem

From a research or social science perspective, the results of evaluations are often vague and uncertain, and evaluations rarely can make confident claims backed by solid evidence and research methodologies. Often, the main concern with evaluation results is not negative evaluations that find problems with peacebuilding programs; it is that evaluations do not reach any strong conclusions, positive or negative.

The main concern with evaluation results . . . is that evaluations do not reach any strong conclusions, positive or negative.

This raises complicated questions about what counts as solid evidence and methodology. The second meeting of the PEP project addressed the question of what different audiences, from the field level to the governmental funder level, would consider good evidence. Not surprisingly, the answers are quite different. For instance, field-level staff need a more detailed anthropological approach. Government officials want a more quantitative, data-driven approach. Neither of these standards should be considered superior to the other. However, both audiences, by their own standards, would find the results of most evaluations lacking.

The Accountability Chain Problem

In most peacebuilding projects there is a chain of oversight from the activity level to the level at which funds are approved. This chain might run from a community-based organization through a national-level NGO, international NGO (INGO), and donor organization to the ultimate funder, often a legislative body or board of directors. Each level in the chain is tasked with holding the level below it accountable.

This accountability chain has several interrelated problems. First, each level in the chain lacks the incentive to hold the level directly below it accountable; incentives for true accountability require at least one degree of separation on the chain. Imagine a national-level NGO asking for information about a workshop that a community-based organization ran. The NGO's desire to maintain its partnership with the INGO overseeing the broader initiative creates an incentive for the NGO to receive positive news about the workshop and report it to the INGO—so-called happy talk—and, simultaneously, a disincentive for transparency and open sharing of results and lessons learned. The INGO or the funder might have an interest in actually knowing whether the workshop was a success or failure, but normally reporting only happens one step up the chain.

Second, as one moves upward along the chain, actors become less knowledgeable and farther removed from the situation in the field. This means that it becomes more difficult at each level to assess the accuracy and quality of the information being passed up the chain. Even if higher levels on the chain have incentives to provide proper oversight, they may not have enough knowledge to make informed decisions about what data to gather or what the data collected actually means.

Third, and related to the weak results problem, each level of the chain has different incentives regarding the kind of information it receives. This creates inherent capacity problems as higher levels on the chain demand types of evidence that the lower levels have not been organized to regularly produce as part of their own work. Of particular importance, higher levels on the chain, because they are likely overseeing many projects, need to receive distilled, aggregated, comparable information. So, for instance, a field-level program manager may be very good at detailed reporting on how political leaders received peacebuilding activities in a given area; this information is likely crucial to guide the implementation of future activities within the project, but provides little information that can be aggregated and compared across programs in different countries.

The RFP Problem

RFP is being used here as shorthand for the dominant strategies by which funders distribute funds to implementers. It is a familiar and ubiquitous process, moving from the strategic planning of the funder to the release of the RFP to the development and review of proposals to the selection of awardees and implementation of projects. This process as it is often implemented undermines good evaluation practice in a number of ways.

The RFP problem is a large and complicated issue, fundamental to almost everything that occurs in the field. To focus the discussion, this report addresses three consequences of the RFP problem that figured prominently in the PEP meeting discussions. First, the RFP system hinders collaboration and coordination among both funders and implementers. Most funders issue RFPs independently of one another; they struggle to release joint RFPs or synchronize them toward a broader goal. They also often have difficulty with evaluation reports that assess the results of multiple projects or the work of multiple grantees. Meanwhile, at the implementer level, most RFPs award funding for projects to single grantees. Even when awardees are consortiums, normally one organization is ultimately responsible for implementing the project. Partnerships might emerge for given projects, but the system makes more sustained and systematic collaboration difficult.

Second, because the RFP process is a competitive winner-take-all process, there is an incentive for organizations to overpromise in their proposals. Once the project is initiated, there is a subsequent disincentive to effectively evaluate because the project will likely not achieve the overly ambitious goals laid out in the proposal.

The NGO's desire to maintain its partnership with the INGO overseeing the broader initiative creates . . . a disincentive for transparency and open sharing of results and lessons learned.

Third, the RFP system hinders the integration of learning into proposals, implementation of projects, and into subsequent RFPs. Funders are often under pressure to spend money, hampering their ability to conduct a full analysis, implement a strategic planning process, or integrate the learning from previous evaluations and assessments into a current RFP. For implementers, timelines to submit proposals are often short, hindering implementers' abilities to integrate learning at the project design stage or to design projects that allow for effective evaluation in the future. As one PEP participant put it, "somewhere, somebody writes a proposal, but never actually reads the report from the ground." Note the important point, a report was written. The problem is not the production of information, but the utilization of that information.

Vicious Circles

The internal logic of the above four problems and the linkages among them create several vicious circles. To begin, in some cases, there are mutually reinforcing dynamics within just one of the problems. Within the accountability chain problem, the lack of incentives for an organization to report honestly creates a lack of incentives for others as well. Just as people do not like to pay taxes if they think everyone else is avoiding them, no one wants to be the only organization honestly and transparently reporting on results. Second, the weak results of evaluations reinforce the tendency for weak results in future evaluations. Participants noted at the PEP meetings that when an organization decides to undertake an evaluation, it looks to existing evaluations for models. The current standard for a project-level evaluation is to conduct a document review and a short field visit, followed by writing an evaluation report. This process can be quite useful, but it is difficult for it to produce strong conclusions. Thus, weak evaluations create additional weak evaluations.

In some cases, two of the problems are mutually reinforcing. For instance, the history of the scale problem exacerbates the RFP problem. In the past, most funders have focused on funding individual projects, which has created processes and a knowledge base within both funders and implementers that deal with specific projects. Donor organizations often do not understand how to coordinate their efforts effectively, or support complex, interlocking initiatives with multiple funders and implementers. In many cases, this lack of understanding has hindered the success of complex efforts, causing funders to retreat to project-based models.[21]

The accountability chain problem and the weak results problem are also mutually reinforcing. The weak results of most evaluations do not engender genuine accountability and learning. Meanwhile, a negative evaluation can still damage an organization. Together, these problems create a situation in which most organizations have very little incentive to make honest assessments of their work. Instead, most look to muddle through, reporting weakly positive results up the chain and understanding that not much will come about as a result.

Most look to muddle through, reporting weakly positive results up the chain and understanding that not much will come about as a result.

More complex loops involve several of the problems. For example, the ultimate funders of projects, often legislative bodies, pressure funders to provide strong oversight and achieve measurable, attributable results with the funding that is provided. These contribute to an RFP system that funds individual projects, which can be more easily monitored. This creates the scale problem, which in turn creates weak results, because meaningful peace outcomes often do not emerge from the work of one project. The weak results travel up the accountability chain from donor organizations to the ultimate funders—who, because the results are weak, place additional pressure on funders to create RFPs for initiatives that can be closely monitored and provide attributable results.[22] Similar dynamics can be caused by projects that overpromise and are therefore judged as unsuccessful.

Second, accountability chain problems create weak evaluation results, because of the incentives problem and capacity problems, whereby each level of the chain normally can-

not provide the types of results that higher levels require.[23] This reinforces the problem with RFP systems that are not designed to integrate learning in a credible, systematic way. Because funders do not receive evaluations that provide strong, clear guidance for future projects, other incentives within the RFP system—for instance, the need to move money quickly, or geopolitical concerns—win out over designing systems that allow the systematic integration of previous learning. The lack of systems within funders that respond effectively to solid evaluation results undermines incentives for implementers to establish strong learning systems that would create these stronger results.

The examples provided here are not the only vicious circles that an observer could identify; the complexities of these problems create a variety of negative feedback loops. The common element of these feedback loops, their self-perpetuating nature, points to the need for whole-of-field solutions. Rather than attempting to address the shortcomings of peacebuilding evaluation simply through individual project evaluations, the actors engaged in peacebuilding evaluation need to shift their perspective so as to understand the linkages among the various elements of the evaluation ecosystem. This does not imply a major overhaul of the peacebuilding field as much as small but important changes in how peacebuilding evaluation is approached so as to undermine these negative and self-perpetuating dynamics.

Whole-of-Field Solutions and the Peacebuilding Evaluation Project

The vicious circles described above are what Michael Masuch defines as both "self-correcting" and "deviation amplifying," meaning that they create and reinforce stagnation.[24] This makes the dynamic often more difficult to address because it does not create a crisis in the same way as "self-reinforcing" vicious cycles, such as bank runs, do. Because of the dynamics' self-correcting nature, a single organization can affect them only with great difficulty, making a more systemic approach necessary. PEP participants seem to have intuitively understood this, as many of the solutions they offered during the meeting sought to address these problems at a systemic level.

As mentioned above, however, working at a systemic level does not suggest massive, large-scale interventions. It has been shown that the right kind of small intervention can successfully interrupt the dynamics that create vicious circles.[25] Many of the interventions described below are small-scale and focused, but all seek to affect the field at a broader, systemic level, by strengthening norms, building consensus, and disrupting practice and creating alternatives.[26]

Each of these categories of solutions is defined by large, complex processes, and by comparison, some of the solutions discussed may seem small-bore, even pedestrian. This is the case because PEP participants were encouraged to speak from their experiences. The participants arrived at practical solutions that they can envision actually being implemented, but because the solutions focus on systemic issues, they have the potential to spark further progress in the field. In each category, initiatives have already begun—either initiated by organizations previously and discussed during the PEP meetings or arising out of the PEP process itself. Those mentioned here, however, are not meant to be a comprehensive list of the many initiatives under way throughout the field as a whole.

Strengthening Norms

Many of the PEP participants' solutions sought to strengthen norms within the peacebuilding field—norms that would create expectations, perhaps even moral obligations, for both organizations and individuals regarding the implementation of effective evaluation practices.

Through the course of the discussion, two related norms emerged that were of particular importance. The first is the norm of honest and transparent assessment of programming. The second, related norm is that organizations have a responsibility to continually learn from their work in order to improve their practice.

Both of these norms are related to the more fundamental ideas of upward and downward accountability, namely, that organizations should assess and learn from their efforts as a means of holding themselves accountable to their funders and their beneficiaries. Without such assessment and learning, organizations may harm those they are trying to help and waste funders' money.

PEP participants identified several ways to work directly on the above norms. The first involved creating a professional mentoring and support network for organization leaders on evaluation issues. One participant said this would be far more valuable than additional written reports or documents, illustrating the importance of relationships and social interaction over further technical knowledge. A second suggestion was to conduct peer-review sessions, in which other members in the field reviewed evaluation documents, including terms of reference and evaluation reports. These meetings could be part of a broader set of meetings in which individuals brought key evaluation challenges to a peer group for discussion and problem solving. Third, fieldwide transparency awards or seals of approval could be created, acknowledging organizations committed to transparency regarding evaluation. Fourth, the peacebuilding field could increase its participation in the American Evaluation Association (AEA) conference and other activities.[27] This would create more awareness, for instance, of the set of guiding principles the AEA has articulated.[28] Finally, donors could work toward a standards document, addressing issues of transparency, the percentage of budget that should go to evaluation, and other process issues, such as allowing time for conflict analysis and establishing baselines. These standards could also require the contribution of evaluation products, such as data sets or interview protocols, into a fieldwide commons.

Norms and the Vicious Circles Joseph P. DeMarco defines a norm as a "rule that regulates voluntary behavior."[29] Applied to the members of a community, the norm creates expectations for what is proper behavior.[30] One can see this dynamic in the various strategies PEP participants identified to strengthen norms related to peacebuilding evaluation. Publicly recognizing transparent organizations can help strengthen the norm of transparency and change other organizations' understanding of what is expected of them in regard to sharing their evaluation results publicly. Participating in a wider community of practitioners, such as the AEA, which already has a strong framework of norms in place regarding how to conduct honest assessments, can help increase the strength of these norms within the peacebuilding field. Similarly, creating common standards among funders regarding evaluation are almost by definition an attempt to establish expectations regarding proper practice.

Strengthening norms can affect the vicious circles described above in several important ways. First, altering expectations about the behaviors of others can undermine the accountability chain problem. As mentioned above, no one wants to report results honestly and transparently if one believes no one else is doing so. An expectation that others are being honest, however, changes the calculation. Just as an organization does not want to be among the first to report honestly, it also does not want to be among the last.

Second, strengthening norms regarding learning and improving practice can help undermine the mutually reinforcing dynamic between the accountability chain problem and the weak results problem. Building stronger learning systems within an organization because of strengthened norms can create a demand for stronger results down the accountability chain. One of the organizations that participated in the PEP process has committed to conducting more holistic assessments of their core programmatic strategies.[31] To do this, they are working intensively with their field offices to improve their ability to conduct assess-

Just as an organization does not want to be among the first to report honestly, it also does not want to be among the last.

ments of their practice at that level. This will improve the results that are passed up to the headquarters level, which should, in turn, strengthen the learning system headquarters is establishing because it has better inputs.

Third, the same dynamic can alleviate the more complex problem described above, in which weak results and accountability chain problems create RFP processes that do not effectively integrate learning and best practices. As in the example above, norms of learning and honest assessment can create stronger learning systems and stronger evaluation results for both funders and implementers. Norms of transparency can also create more awareness about what has been learned, making it harder for funders to continue to implement RFP processes that do not integrate the new lessons. One can see this dynamic in the recent strategic reorientation of the Department for International Development-UK (DFID), which focuses on delivering results and adding value. While funders often speak in these terms, DFID actually made difficult decisions regarding eliminating certain country programs and refusing to fund organizations that did not demonstrate their effectiveness in the past.[32] As of this writing, it is too soon to see how this reorientation will affect DFID RFP processes.

What Is Already Happening Following their own suggestion, PEP participants created panel proposals that have been submitted to the AEA's 2011 annual meeting. The Alliance for Peacebuilding will arrange a series of peer-review and community-of-practice meetings over the next year. Search for Common Ground and American University are launching a learning portal for design, monitoring, and evaluation of peacebuilding. A key goal of the portal, according to the designers, is to help shift norms regarding transparency and sharing of evaluation results through, for instance, drawing positive attention to those organizations that agree to post their evaluations on the portal.[33]

Building Consensus

Another set of solutions suggested by PEP participants can be grouped under the heading of building consensus on effective peacebuilding practices and evaluation practices. Suggestions for interventions of this type included holding an evidence summit to articulate, based on evidence, what the field knows. This would help shift the field toward an evidence-based approach and away from a "gut-based" approach. Related to the evidence summit was the suggestion that the field create what was termed a "received wisdom" document, containing a number of things known to be good practice as well as bad practice. Funders and implementers could also initiate a process of working toward agreement on a small number—perhaps four or five—of core peacebuilding outcomes, such as reducing the number of violent deaths in a certain area, by which to measure a project's success. Similarly, evaluation reports could be synthesized annually to distill what the field has learned over the course of the year. These syntheses could be further distilled into two- to three-page page pamphlets on key best and worst practices.

PEP participants also discussed creating space for and supporting peacebuilding bloggers. In the international development field, the blogosphere has been one of the most powerful mechanisms for articulating a consensus and holding organizations accountable to it. Blogs such as AidWatch,[34] Chris Blattman, the Center for Global Development, Owen Abroad, and many others have become vocal and consistent critics of substandard aid practices. Of course, the consensus on what constitutes good aid practice is constantly contested and evolving. Nonetheless, the standards are being articulated and enforced, if only through moral suasion. The peacebuilding field—with a few exceptions, such as some of Chris Blattman's work—does not have a similar conversation under way.

Consensus Building and the Vicious Circles During the PEP meetings and in other evaluation conversations, there has been a strong focus on the importance of evaluation

Funders and implementers could also initiate a process of working toward agreement on a small number of core peacebuilding outcomes by which to measure a project's success.

for learning. But for a focus on learning to be meaningful, eventually the field must be able to articulate clearly and confidently a small number of things the field has learned—while acknowledging they may change over time—and hold organizations accountable if their practices run counter to these findings. Building this kind of consensus regarding effective and ineffective practices can undermine the vicious circles described above in several important ways; for example, consensus on a small number of core outcomes or basic effective practices simplifies the reporting up the accountability chain and improves intelligibility. This disrupts the self-perpetuating dynamic between weak results and the accountability chain. To understand why this is the case, it is helpful to look at the World Bank's 2011 World Development Report (WDR).[35]

The WDR took a large amount of research and distilled it down to a simple conclusion, namely, that three things matter in preventing a reoccurrence of conflict: security, justice, and jobs.[36] This claim will be contested, but if consensus emerges that it is valid, it will allow reporting up the accountability chain to focus on these three things. Security, justice, and jobs are all complex phenomena, so this does not solve the problem, but it does provide a start. It creates an expectation that a strong evaluation result will be able to make a claim about how a project contributed to one of these three categories. Compare this to a situation, which is common now, where those on different links of the accountability chain do not even agree on what results are important.

Building consensus also helps undermine the scale problem, which in turn helps undermine the more complex dynamics in which the scale problem, the RFP problem, and the accountability chain problem interact. As was discussed above, the scale and weak results problems often lead to RFPs focused on individual projects, which in turn re-create narrowly focused initiatives and weak results. Creating consensus on a small number of effective practices can help undermine this situation. Again, using the WDR as an example, the focus on security, justice, and jobs makes it easier to link micro-level results—such as increased perceptions of security in a community—with macro-level outcomes, as well as to aggregate results across multiple projects in order to report up the accountability chain, including to ultimate funders, such as legislative bodies. This can undermine both the scale problem and the problem of overpromising in proposals. Similarly, a focus within the accountability chain on a smaller number of important results can feed more effectively into future RFP processes, making it easier for funders to integrate learning and past results into their RFPs. This in turn incentivizes implementers to build capacity and develop stronger learning systems.

The goal of a consensus-building process is not to agree on a small number of permanent, universal truths. As is noted above, the WDR conclusions will likely be contested, but in many ways that is the point. They have clearly articulated a focused set of conclusions regarding effective peacebuilding.[37] Instead of finding universal truths, the goal is to articulate contestable claims about effective and ineffective peacebuilding practices, as well as clear statements about what we still do not know. The strategies described above, including holding an evidence summit, developing evaluation syntheses, creating a process to build consensus on a small number of core peacebuilding outcomes, and supporting peacebuilding bloggers can all be harnessed toward this goal.

What Is Already Happening USIP and the Alliance for Peacebuilding have begun work on an evidence summit to be held later in 2011. Discussion about the evidence summit took place before the release of the WDR, which will likely change the nature of the event. The learning portal described above will seek to synthesize evaluations that have been submitted to the portal into various types of summaries to capture what was learned. Finally, International Alert provides a useful example of the importance of a consensus-building process at the organizational level. Phil Vernon of International Alert gave a presentation at the third PEP meeting on their ongoing initiative to improve evaluation practice. Interestingly,

the first step in their process was one of consensus building and articulating in writing their core beliefs about effective peacebuilding. Much of the process International Alert went through at an organizational level is relevant at a fieldwide level.[38]

Disrupting Practice and Creating Alternatives

Another group of PEP participants' suggestions focused on the need to disrupt current practices and provide alternatives to them. As noted above, the vicious circles holding back progress in peacebuilding evaluation cause unwanted stagnation. In this situation, disrupting normal practice and demonstrating the viability of alternatives can be beneficial. Individuals and individual organizations have some autonomy within the larger structural constraints the field imposes; the suggested actions below can be taken because of that autonomy, but at the same time, they can affect those broader structural constraints.

First, as a means of addressing the scale problem, a group of implementers could initiate a demonstration project that creates collaborative monitoring and evaluation mechanisms in a certain sector, or on a certain conflict. A similar project could demonstrate real-time or continuous monitoring of a project. One critique raised during the meetings was that evaluations often memorialize past practices. If there is to be a focus on learning, then evaluation needs to create more continuous feedback loops during the course of the project. The goal here is to show how a better synthesis between monitoring and evaluation can be achieved.

Second, implementers in collaboration with funders could create failure stories or worst-project awards to disrupt standard practices regarding success stories, best practices, and "happy talk." One PEP participant noted that in many areas a 50 percent success rate is still very good and that the field should not hold itself to a 100 percent success standard. Instead, failures should be identified to increase the capacity to learn from them.

Third, a blog space could be developed, perhaps as part of the learning portal, that allows program managers and community members in the field to provide their perspectives. The contributions could be curated by an administrator at the portal to ensure anonymity if desired. More direct, honest assessments of how evaluation is practiced at the field level could help disrupt more harmful practices at the headquarters and funder levels. If headquarters and donor organizations began soliciting this input, it would also disrupt the standard reporting practices in the accountability chain in which organizations only report to the next level up.

Disruption, Alternatives, and the Vicious Circles

The above suggestions can affect the vicious circles in several concrete ways. First, opening up space for failure stories and creating mechanisms through which those at the field level can speak to the peacebuilding field as a whole on evaluation issues can disrupt the accountability chain–weak results dynamic. An increased willingness to acknowledge failures, combined with a more open multilateral conversation among funders, headquarters, the field level, and even the community level can foster more continuous learning processes among these levels, as opposed to the more formal, unilateral, and invariably cautious reporting that is common now. GlobalGiving's Storyteller Project and Mobile Accord's GeoPoll project are interesting early experiments at creating this type of ongoing, multilateral conversation.[39]

Second, as is argued above, the history of weak evaluations and the history of the scale problem both create vicious circles. Evaluations that have created weak results in the past are used as models for future evaluations. Regarding the scale problem, funders and larger implementers do not have a good understanding of how to fund, manage, or evaluate

Implementers in collaboration with funders could create failure stories or worst-project awards to disrupt standard practices regarding success stories, best practices, and "happy talk."

broader, complex initiatives. Demonstration projects that create examples of strong evaluations or show concretely how collaboration on evaluation can contribute to overcoming the scale problem can undermine the self-reproducing nature of these problems.[40]

Demonstration projects that illustrate how to produce stronger evaluation results—particularly for broader, more complex programs or a group of programs working on a specific conflict or sector—can also affect the more complex dynamics, described above, that cause funders to retreat to funding individual projects that are measurable and attributable. These projects create weak results because of the scale problem, which creates renewed pressure from the funders. Demonstrating clearly how broader, interlocking initiatives can be evaluated successfully can reverse this dynamic and make funders more amenable to funding more complex programming models.

What Is Already Happening

A group was formed at the final PEP meeting to collaborate on a demonstration project designed to show how evaluation collaboration might work. The organizations—currently CARE, Partners for Democratic Change, and Search for Common Ground—have agreed to synchronize their evaluation processes for their programs dealing with women's empowerment and leadership development. Other organizations may join the initiative.

Meanwhile, at the aforementioned evidence summit, there will be a focus on practices that do not work. The goal will be to contribute to a consensus on a small number of practices that don't work and develop means to hold organizations accountable if they employ such practices. It is not clear yet if those working in the peacebuilding field are willing and able to report openly on their own individual failures; this will be assessed.[41]

Conclusions and Recommendations

This report has identified many of the specific next steps that should be and are being taken. Three more general lessons emerged from the PEP discussions, not so much regarding what actions need to be taken, but how the field should orient itself toward the problem of making progress in peacebuilding evaluation.

First, improving the evaluation of peacebuilding work involves fostering organizational change and restructuring the relationships among organizations, as opposed to simply overcoming a technical, research, or methodological challenge. Even very solid evaluation research, unless it is targeted to address specific problems, such as the demonstration projects described above, can rarely impact the larger systemic problems in the field. In contrast, just good evaluation research that takes place in a context of strong learning systems and strong norms of transparency can create demand going forward for better evaluation research and better methodologies.

Second, when a simple vicious circle between two elements is diagrammed, there is always an arrow and a return arrow. In bringing funders and implementers together, the PEP process illustrated the crucial lesson of understanding the return arrow. Progress will be made on peacebuilding evaluation by crafting solutions that actually disrupt vicious circles, as opposed to simply demanding that organizations, whether funders or implementers, do things differently. For instance, implementers commonly critique funders for working on too short a time scale, while funders demand more rigorous data from implementers. Simply asserting things in this way is not helpful, as it does not acknowledge the dynamics that are giving rise to the problem. As one PEP participant noted, "All these dysfunctional behaviors have incentives behind them."

Improving the evaluation of peacebuilding work involves fostering organizational change and restructuring the relationships among organizations.

Third, it is important to learn from peer fields. It is clear that the field of international development is ahead of the peacebuilding field in regard to evaluation. This is not only an issue of stronger methods or research, which might be expected, given the size of development projects compared with most peacebuilding projects and the fact that development activities, all else being equal, are more amenable to rigorous evaluation methodologies. The international development field is also ahead of peacebuilding in strengthening norms, such as increasing transparency by publishing evaluation reports;[42] building consensus through the willingness to articulate standards for bad practices and hold organizations accountable to those standards;[43] and creating alternatives, through, for instance, the work of the aforementioned Poverty Action Lab. Development practitioners would be quick to say that none of the mechanisms devised in their field is perfect, and there is much work yet to be done. However, the peacebuilding field can learn much from their progress.

The PEP process was itself a contribution to the whole-of field approach articulated in this report. Bringing implementers and funders together was intended to contribute to the fieldwide changes that this report argues are necessary to improve the practice of peacebuilding evaluation. The process succeeded in strengthening existing avenues and opening new ones for facilitating changes. However, for progress to continue, these changes must become embedded within evaluation processes themselves. When implementing evaluations, it is important to reflect both on how the evaluation process can succeed and how it can affect the field more broadly through strengthening norms, building consensus, and creating alternatives. As the analysis above makes clear, small steps within an evaluation process can create progress in peacebuilding evaluation, if these steps are consciously undertaken to positively affect systemic, whole-of-field dynamics.

Notes

1. The first meeting dealt with mapping the field, discussing the general challenges and opportunities for improving peacebuilding evaluation. The second covered evaluation methodology, evidence, and the various audiences for evaluation results. The third discussed change at the organizational level. The fourth discussed changing the field itself.

2. Meetings were held under Chatham House rule, although participation was openly acknowledged. Nothing in this report is attributed to individual participants. For a full list of participants, see Alliance for Peacebuilding, *Starting on the Same Page: A Lessons Report on the Peacebuilding Evaluation Project* (Washington, DC: Alliance for Peacebuilding, 2011), http://afpb.site-ym.com/resource/resmgr/Docs/PEP_Lessons_Report_0521201_A.pdf.

3. While this report is informed by the discussions at the PEP meetings, it should not be considered direct reporting on those conversations. Instead, the report attempts to place those conversations in a broader context regarding what needs to be done to improve peacebuilding evaluation. The conclusions reached here are the responsibility of the author, not the PEP participants. In addition to this report, the Alliance for Peacebuilding (AFP) recently released *Starting on the Same Page.*

4. This definition combines what the OECD refers to as peacebuilding and conflict prevention. See Organization for Economic Cooperation and Development, "Guidance on Evaluating Conflict Prevention and Peacebuilding Activities," http://www.oecd.org/secure/pdfDocument/0,2834,en_21571361_34047972_39774574_1_1_1_1,00.pdf. See also "Peacebuilding" in Dan Snodderly, ed., *Peace Terms: Glossary of Terms for Conflict Management and Peacebuilding* (Washington, DC: Academy for International Conflict Management and Peacebuilding and United States Institute of Peace, 2011), http://www.usip.org/peace_terms.pdf.

5. Organization for Economic Cooperation and Development (OECD), "Evaluating Conflict Prevention and Peacebuilding Activities," Factsheet 2008, 2, http://www.oecd.org/dataoecd/36/20/39289596.pdf.

6. Cheyanne Church and Mark Rogers, *Designing for Results: Integrating Monitoring and Evaluation in Conflict Transformation Programs* (Washington, DC: Search for Common Ground, 2006), 93.

7. John Paul Lederach, Reina Neufeldt, and Hal Culbertson, *Reflective Peacebuilding: A Planning, Monitoring, and Learning Toolkit* (South Bend: Joan B. Kroc Institute and Catholic Relief Services, 2007), 1.

8. Cheyanne A. Church, "Peacebuilding Evaluation: From Infancy to Teenager," *New Routes* 13, no. 3 (2008): 4.

9. Church, "Peacebuilding Evaluation," 4.

10. There are many reasons for this, a discussion of which is beyond the scope of this report. One indicator of these pressures, however, is the high level of demand to participate in the Peacebuilding Evaluation Project itself. Even though the initiative required a commitment to participate in four meetings and there was no stipend or honorarium to do so, forty-four members of the Alliance for Peacebuilding applied. In the end, fourteen were chosen based on a set of criteria development by AFP.

11. Much of this pressure comes from the international development field, where a range of critics are pressuring organizations to do more rigorous evaluations. For one example, see the critique of the Millennium Villages project for not doing more rigorous evaluation in Michael Clemens and Gabriel Demombynes, "When Does Rigorous Impact Evaluation Make a Difference? The Case of the Millennium Villages," Center for Global Development working paper no. 225, Washington, DC. See also Jonathan Starr, "The Business of International Aid," *Wall Street Journal,* April 11, 2011, http://online.wsj.com/article/SB10001424052748704425804576220524034207558.html?mod=googlenews_wsj; or the dialogue on World Vision's gifts in kind program that took place on the blog AidWatch, http://aidwatchers.com/2011/03/world-vision-super-bowl-shirts-the-final-chapter/.

12. In addition to pressure from funders, Mary Anderson's work is important here. There is now consensus in the field that well-meaning, even well-implemented programs, can do harm. This creates additional pressure for proper evaluation to avoid these harmful outcomes. See Mary Anderson, *Do No Harm: How Aid Support Peace—Or War* (Boulder, CO: Lynne-Rienner, 1999).

13. Peacebuilding-specific works include Catholic Relief Services, *Gain Peacebuilding Indicators* (Baltimore: Catholic Relief Services, 2010); Susan Allen Nan, *Theories of Change and Indicator Development in Conflict Management and Mitigation* (Washington, DC: US Agency for International Development, 2010), http://pdf.usaid.gov/pdf_docs/PNADS460.pdf; OECD, "Evaluating Conflict Prevention"; Lederach, Neufeldt, and Culbertson, *Reflective Peacebuilding;* Church and Rogers, Designing for Results; Marc Howard Ross and Jay Rothman, eds., *Theory and Practice in Ethnic Conflict Management: Conceptualizing Success and Failure* (Basingstoke and London: Macmillan, 1999).

14. See Sarah Earl, Fred Carden, and Terry Smutylo, *Outcome Mapping: Building Learning and Reflection into Development Programs* (Ottawa: International Development Research Centre, 2002).

15. See Jessica Dart and Rick Davies, "A Dialogical, Story-Based Evaluation Tool: The Most Significant Change Technique," *American Journal of Evaluation* 24, no. 2 (2003): 137–155.

16. For further information see http://www.povertyactionlab.org/. The Innovations for Poverty Action unit of the Poverty Action Lab recently launched a Post Conflict Recovery and Fragile States program with the goal of applying RCT methodologies to postconflict programming.

17. For a more complex articulation of systemic dynamics in the field, which has significant overlap to the problems laid out here, see the systems map developed by PEP participants which is included as Annex 1 in Alliance for Peacebuilding, *Starting on the Same Page.*

18. CDA Collaborative Learning Projects is generally credited with popularizing the term *peace writ large.* For a good discussion, see Diana Chigas and Peter Woodrow, "Envisioning and Pursuing Peace Writ Large," Berghof Research Center for Constructive Conflict Management, http://www.berghof-handbook.net/documents/publications/dialogue7_chigwood_comm.pdf.

19. For a discussion of this problem as it impacts the larger social change field, see John Kania and Mark Kramer, "Collective Impact," *Stanford Social Innovation Review* (Winter 2011), http://www.ssireview.org/articles/entry/collective_impact/.

20. For two examples, see Maria Hadjipavlou and Bülent Kanol, *Cumulative Impact Case Study: The Impacts of Peacebuilding Work on the Cyprus Conflict* (Cambridge, MA: CDA Collaborative Learning Projects, 2008), http://www.cdainc.com/cdawww/pdf/casestudy/rpp_cyprus_cumulative20case20final_20806031_Pdf_1.pdf, and Janet Murdock and Alfiado Zunguza, *Cumulative Impact Case Study: The Cumulative Impacts of Peacebuilding in Mozambique* (Cambridge, MA: CDA Collaborative Learning Projects, 2010), http://www.cdainc.com/cdawww/pdf/casestudy/rpp_cumulative_cases_mozambique_final_Pdf.pdf Mozambique.

21. A recent prominent example is the Multi-Donor Trust Fund (MDTF) in Sudan. An evaluation of peacebuilding in Sudan indicated that donors who worked bilaterally were more effective than pooled funding mechanisms such

as the MDTF. Netherlands Ministry of Foreign Affairs, *Aiding the Peace: A Multi-Donor Evaluation of Support to Conflict Prevention and Peacebuilding Activities in Southern Sudan 2005–2010* (The Hague: Netherlands Ministry of Foreign Affairs, 2011), 37, 57, 71–74.

22. For U.S. organizations, part of this pressure comes from an increased focus on compliance issues. See Andrew Natsios, "The Clash of the Counter-Bureaucracy and Development," Center for Global Development essay, July 2010.

23. This problem is mitigated but not solved by using an external evaluation consultant. Consultants bring more capacity down the chain, for instance, from headquarters level to the field, but they still depend on those in the field for various kinds of support and inputs, such as previous data collection efforts.

24. Michael Masuch, "Vicious Circles in Organizations," *Administrative Science Quarterly* 30, no. 1 (1985): 16.

25. This, of course, is the argument made famous in Malcolm Gladwell, *The Tipping Point: How Little Things Can Make a Big Difference* (Boston: Little, Brown, 2000). The basic point is that situations characterized by nonlinear dynamics, such as a vicious circles, are subject to tipping points, cascades, and other types of self-perpetuating phenomena that can create large-scale change. For a famous argument of this type related to street protests, see Timur Kuran, "Sparks and Prairie Fires: A Theory of Unanticipated Political Revolution," *Public Choice* 61, no. 1 (1989): 41–74.

26. These categories should be seen as fuzzy and overlapping. Many of the solutions offered have effects that could fit in more than one category.

27. This was suggested by PEP participants, but it is not new. See Church, "Peacebuilding Evaluation," 6.

28. American Evaluation Association, "American Evaluation Association, Guiding Principles for Evaluators," http://www.eval.org/publications/GuidingPrinciplesPrintable.asp.

29. Joseph P. DeMarco, "Norms: Their Moral Status," http://www.lawandbioethics.com/demo/Main/EthicsResources/Moral_status_of_norms.htm; see also Matthew Interis, "On Norms: A Typology with Discussion," *American Journal of Economics and Sociology* 70, no. 2 (April 2011): 424–38, http://onlinelibrary.wiley.com/doi/10.1111/j.1536-7150.2011.00778.x/abstract.

30. Interis, "On Norms," 428.

31. Interestingly, this commitment was donor-supported, but not donor-driven. Identifying the effects of norms is notoriously difficult, but in this case the strengthening of the norms of learning and honest assessment in the peacebuilding field appear to have played at least a part in the organizational decision to improve the organization's learning systems.

32. For a summary, see DFID, "The Future of UK Aid," press release, http://ht.ly/45ysx.

33. USIP recently awarded a grant to support this project as part of its 2010 annual grant competition.

34. AidWatch was a key player in this process, but unfortunately was closed by its organizers on May 18, 2011.

35. World Bank, *World Development Report 2011: Conflict Security and Development* (Washington DC: World Bank, 2011). The report and the input papers are available at http://wdr2011.worldbank.org/.

36. The World Development Report makes the additional claim that the core challenge for the peacebuilding field is now preventing the reoccurrence of conflict, as opposed to the emergence of new conflicts. World Bank, *World Development Report,* 57–58.

37. It is illustrative to compare the 2011 World Development Report to the U.S. Agency for International Development document, "Conducting a Conflict Assessment: A Framework for Analysis and Program Development," Washington, DC, 2004, http://www.usaid.gov/our_work/cross-cutting_programs/private_voluntary_cooperation/conf_conflict_assessment.pdf. This document has a table entitled "Causes of Conflict: Overview," which lists twenty-three causes of conflict but contains nothing that could be construed as a contestable claim.

38. For one outcome of that process, see International Alert, *Programming Framework for International Alert: Design, Monitoring, and Evaluation* (London: International Alert, January 2010), http://www.international-alert.org/sites/default/files/library/012010ProgrammingFrameworkForAlert.pdf.

39. For information on the Global Giving Project, see Britt Lake, "Collecting Beneficiary Feedback—GlobalGiving's Storytelling Project," Global Giving, December 3, 2010, http://tools.blog.globalgiving.org/2010/12/03/collecting-beneficiary-feedback-globalgivings-storytelling-project/; for information on GeoPoll, see Mobile Accord, "GeoPoll," http://www.mobileaccord.com/services/smsPolling.aspx.

40. The set of randomized control trial (RCT) evaluations published by the MIT Poverty Action Lab contributed to this dynamic within the international development field. The important issue here is not the RCT methodology, but the demonstration of how evaluations that produce stronger results could be implemented. This changed the conversation within the development evaluation community. As noted above, they have now launched a program focused on postconflict programming. It it too soon to know what effect this will have on the peacebuilding field.

41. For an example of failed reporting in the information and communications technology for development (ICT4D) sector, see anoushrima, "How to Roll Your Own FAILFaire," MobileActive.org, http://www.mobileactive.org/roll-your-own-failfaire.

42. It is possible to find numerous development evaluations online. Similar internet searches reveal very few peacebuilding evaluations.

43. A more formal manifestation of this is the standards articulated in InterAction, "PVO Standards Interpretive Guidance and Member Guidelines," http://www.interaction.org/document/pvo-standards-interpretive-guidance-and-member-guidelines. A more informal accountability mechanism is the network of blogs, some backed by organizations such as the Center for Global Development, which are dedicated to oversight of the development sector. See also William Easterly and Claudia R. Williams, "Rhetoric versus Reality: The Best and Worst of Aid Agency Practices," forthcoming in *World Development*. For a discussion, see DFID, "The Future of UK Aid."

Of Related Interest

- *Peace Education: State of the Field and Lessons Learned from USIP Grantmaking* by Mari Fitzduff and Isabella Jean (Peaceworks, August 2011)
- *Gender, Conflict, and Peacebuilding: State of the Field and Lessons Learned from USIP Grantmaking* by Kimberly Theidon and Kelly Phenicie; with Elizabeth Murray (Peaceworks, September 2011)
- *Civic Education and Peacebuilding: Examples from Iraq and Sudan* by Daniel H. Levine and Linda S. Bishai (Special Report, October 2010)
- *Graduate Education and Professional Practice in International Peace and Conflict* by Nike Carstarphen, Craig Zelizer, Robert Harris, and David J. Smith (Special Report, August 2010)
- *Conflict Assessment and Intelligence Analysis: Commonality, Convergence, and Complementarity* by Lawrence Woocher (Special Report, June 2011)
- *Preventing Violent Conflict: Assessing Progress, Meeting Challenges* by Lawrence Woocher (Special Report, September 2009)