Chapter 2

Peacebuilding Evaluation: Two Decades of Evolution

Cheyanne Scharbatke-Church

Conflict and fragility are as significant a global issue today as they were at the turn of the millennium. While the amount of conflict around the globe has ebbed and flowed since the early 1990s, the Council on Foreign Relations 2019 Global Conflict Tracker has more conflicts worsening than improving.[1] The Fragile State Index shows more countries classified in the "Alert" categories in 2018 than were there in 2006, and the Global Peace Index states, "Peacefulness has declined year-on-year for eight of the last ten years. Since 2008, 85 countries have become less peaceful, compared to 75 that have improved."[2]

The number of conflicts is large; the number of people affected are far larger. In 2018, 2 billion people were impacted by conflict and fragility.[3] Looking ahead, it is predicted that "by 2030 two-thirds of the world's poorest will live in a situation of conflict and fragility."[4]

The importance of the issue has not escaped the international community.[5] As the Busan Partnership for Effective Development Cooperation states, "Nowhere are our development goals more urgent than in fragile and conflict-affected states."[6] Peacebuilding's contribution is to catalyze a positive change in these conflicts. To achieve this, the field needs processes to assess effectiveness and offer a means of improvement. Reflective practice, academic research, and indices can all play a role in this, but program monitoring and evaluation (M&E) is by far the most commonly used mechanism in peacebuilding.

This chapter focuses solely on program evaluation—the systematic use of social science data collection methodologies, including participatory approaches, to ascertain the quality and value of a peacebuilding effort (i.e., program or policy). *Quality*, in this case, reviews the strength of the work that was done, and *value* determines if that work made a difference that matters to the people and/or conflict (i.e., relevance). As such, peacebuilding program evaluation is distinct from pure social science research. For instance, research and evaluation have different purposes, timelines, and deliverables (Jean and Ernstorfer 2014). While good evaluation cannot happen without good research skills, good research can happen separate from evaluation. Admittedly, looking exclusively at evaluation is a bit misleading as this suggests that evaluation exists in isolation. In reality, excellent evaluation is associated with quality conflict analysis, strategic program design, useful monitoring, and, ultimately, the program's and organization's ability to learn.

The chapter adopts a "small tent" approach to determine what counts as peacebuilding: it is any practice, policy, or intervention that seeks to address the key driving factors of a conflict, such as discrimination, patronage, or violence. This does not diminish the importance of "big tent" peacebuilding, such as national infrastructure or economic development, it simply makes the distinction between efforts that are directly combating conflict issues and those that are indirectly relevant. This matters to evaluation, because those efforts directly addressing key driving factors of

conflict are typically grounded in complex adaptative systems, which present unique challenges to an evaluation team.

Literature specifically addressing the intersection of these two fields, evaluation and "small tent" peacebuilding (predominantly framed at that time as international conflict resolution), started to appear in the late 1990s. Mark Duffield and Tanya Spencer were two of the first to publish conceptual pieces, in 1997 and 1998, respectively. The practitioner community was then staunchly in a "you can't measure what we do" mindset. In contrast, today practitioners can be forgiven if they feel there are constant demands for "evaluation everywhere all the time." References to evaluation practice can be found in government strategies, international nongovernmental organization (INGO) guidance notes, and every nongovernmental organizations (NGO) website. Looking back at the past two decades of evaluation of peacebuilding practice, this chapter will explore how the field moved from perceived immeasurability to the current state of M&E proliferation and the gaps that remain. The chapter is a study of trends, based on both documented sources and the author's personal observations and reflections.

CHANGES IN PEACEBUILDING EVALUATION PARADIGM AND PRACTICE

The changes to the peacebuilding evaluation field have been vast, spanning the conceptual to the material. Thinking has evolved on what constitutes a result, how change occurs, how we measure change, and how technology can be incorporated into the process.

Emphasis on Results

To understand the changes in evaluation of peacebuilding since the 1990s, we must start by looking at the paradigm transformation experienced in the broader foreign aid field in which peacebuilding is situated. What was

formerly an activity-centric profession focusing primarily on what programs *did*—such as train, mediate, peace journalism—shifted to a (purported) focus on *results*. While the specific meaning of "results" is interpreted differently throughout foreign aid, this new-found emphasis meant that program evaluation started to receive far more attention. This shift also had significant impact on those working on conflict issues, in both positive and negative ways.

The belief that social change programming results were important to understand, and the use of program evaluation as a means to assess them, can both be traced back to domestic public administration programs of the 1960s in the United States, Europe, and Australia (Bush and Duggan 2015). It took almost three decades before these principles were subsequently adopted by the same governments' foreign aid ministries and departments. For example, in "1996, USAID[7] had … established a set of indicators and required country missions to report annually on program results" (Natsios 2010, 23). But it was not until the early 2000s that a series of high-profile global processes on aid effectiveness, punctuated by High-Level Fora of donor and aid-recipient countries, transformed results from a topic to a paradigm that had reverberations through all sectors in the international community.

The first High-Level Forum on Aid Effectiveness took place in Rome in 2003, but the aid effectiveness agenda really came into its own in Paris in 2005.[8] The resulting agreement, the Paris Declaration, laid out five principles to improve official aid, two of which demanded the international community do better on the issue of results. With these two principles officially accepted, determining if and how programming makes a difference was placed squarely on the main agenda of the international community. The concept of effectiveness continued to be emphasized and refined in each of the next High-Level Fora in Accra in 2008 and Busan in 2011.

This was not exclusively a shared agenda among donors and national officials, as civil society also officially launched a parallel process in 2008. The Open Forum for Civil Society Organization (CSO) Development

Effectiveness process resulted in the adoption of the Istanbul Principles.[9] In this declaration, principle eight commits CSOs to realizing positive sustainable change.

As a far younger field than its humanitarian and development cousins, the starting point for peacebuilding actors in the early days of the aid effectiveness agenda was quite different from other sectors. In the mid-to-late 1990s, the American and European peacebuilding agencies were not proactively engaging in results conversations. Maxims such as "you can't measure what we do" and "this work takes generations" were commonly accepted arguments as to why discussions of concrete results were not possible. Results, when addressed at the time, were typically statements of activities accomplished (e.g., thirty women trained in nonviolent conflict resolution). While aspiring visions were common, there was neither a robust or a nuanced conversation regarding what difference a project could be expected to make in a conflict nor much investment in ascertaining if a project got there.

As the global processes progressed, the unique challenges of fragile and conflict-affected states were acknowledged. In the case of the Open Forum on CSO Development Effectiveness, CIVICUS coordinated three thematic consultations with CSOs working in these contexts in the lead up to the development of the Istanbul Principles (Tomlinson 2012). At the official level, a parallel process piloted by the International Dialogue on Peacebuilding and Statebuilding focused on fragility and conflict. This effort resulted in the signing of the New Deal for Engagement in Fragile States by forty states at the 2011 Busan Forum.[10]

Once again, the results theme is apparent, though framed differently than in the High-Level Forum declarations. In the New Deal, there is no single principle or statement vis-à-vis results; rather, the entire declaration is grounded in a "results" orientation. For instance, the five Peacebuilding and Statebuilding Goals lay out the results sought within fragile and conflict-affected states, while the ways of engaging (i.e., Focus) offer several results-centric ideas, such as the commitment to monitoring progress.

The adoption of a results paradigm in the peacebuilding field over the past twenty years has not been without unintended negative consequences. Swisspeace, a practice-oriented peace research institute, has gone so far as to claim "that results-orientation in its currently practiced form is more of a hindrance than a help for achieving better results" (Bachtold, Dittli, and Servaes 2013, 5). No critique of the results paradigm quibbles with the need for better results in their purest form—making a difference that matters in a conflict. The challenges lie in the way the results orientation has been taken from policy to practice predominantly in the donor arena, with the understanding that this has trickle-down impacts on implementing agencies. While there are numerous critiques of the manner in which this results orientation has been operationalized, three of the most significant will be discussed here.

The first concern lies with what counts as a "result." The dominant interpretation is that a result is something that can be counted or observed, rather than a relevant change in the conflict. This interpretation is driven by the desire to simplify, and the perception that it enables measurement. Andrew Natsios, a former USAID administrator, pithily sums up how the desire to measure has influenced the discussion, "Essentially, measurability should not be confused with development significance" (2010, 23).

This push for measurability, coupled with two dominant pressures in donor culture—compliance and short-term perspectives—creates unintended consequences, in effect undermining a true results agenda. In the compliance culture, the work plan that was initially proposed is held up as the standard against which the work done is compared. This creates an unhelpful focus on the activities peacebuilders do, rather than the difference they make. When the short time frames typical to donor processes are then considered, the only "results" generally possible to identify are those that can be counted. The result of this combination of factors is a disproportionate emphasis on quantifiable results, which in peacebuilding commonly boils down to the number of people at an event, or other simplistic figures that do not indicate change in the conflict.

The second concern occurs when working at the scale of most bilateral donors: a scale at which standardization of processes becomes understandably important. Yet it is extremely difficult to develop a standardized process that is responsive to the diverse and challenging contexts (and the various resulting programs) that are found in peacebuilding (see d'Estrée on complexity, this volume). As such, the tendency toward mechanistic and overly simplified processes has further entrenched the results orientation toward things we can count and away from things that may be transformative.

Finally, the emphasis on results has unintentionally created risk aversion in many quarters. Given the high-risk nature of many programs in fragile and conflict-affected states, this has significantly influenced the type of programming that is being done. If a program cannot reliably deliver in the time frames given, the types of results desired—quantitative ones that align to standardization—then are deemed risky. Coupled with the compliance culture of many bilateral donors, there is no incentive to take a risk or determine creative ways of dealing with the complexity that is inevitably involved in changing conflict systems.

Types of Change

Closely related to, and arguably caused by, the shift to emphasize results is the beginning of a transformation in how the field defines results. As the 2011 *United Nations Development Group Results Based Management Handbook* states, "Results are about change." The nascent embracing of *types of change* (e.g., attitudes, social norms, structures, and skills) as the way in which the field needs to think about peacebuilding results marks a significant shift (that aligns with evaluation texts (e.g., Patton 2008). "Change is a difference in the content, form or functioning of something in a context that is distinct from the activity that was being done to catalyse that change" (Scharbatke-Church and Rogers forthcoming). What difference did your work make, not counting the immediate direct products of your work (for example, a training manual, website, or brochure)? This is

the same distinction between *outcomes* and *outputs* found in logic models or other logical frameworks (see also d'Estrée, chapter 1, this volume).

This is a crucial shift from the 1980s and 1990s which emphasized the work peacebuilders did (e.g., dialogues, radio shows, and mediation). It is also distinct from the dominant results framing that came out of the results orientation, which focused upon the outputs created (e.g., number of TV show viewers, number of attendees at a peace festival, or number of brochures disseminated). Twenty years ago, references to some types of change (e.g., attitude and behavior) could be found in work that explained how conflict happens, but this typically did not get translated into program design and evaluation applications (Fisher et al. 2000).

While these concepts are now in the common lexicon of peacebuilding, this does not yet mean that distinctions among *types of change* are always used correctly or well. For example, there is still a strong tendency in the peacebuilding field to promise overly lofty and ill-defined changes, often framed as a program's goal. These aspirational statements, such as "contribute to security" or "build national identity," arguably reference change(s) but are not at the level of specificity that *types of change* offer. This goal inflation is partly due to mythologies around the claims needed for fundraising but also arises from fuzzy, aspirational thinking.

The introduction of *types of change* as applied to peacebuilding is important to evaluation, because it offers potential solutions to two common evaluation challenges: how to evaluate activity lists and how to assess programs against clearly inflated goals. From the late 1990s and well into the 2000s, it was common for evaluators to be tasked with determining if a peacebuilding program was a success or not, based on a list of activities the organization undertook. It is very difficult to ascertain success when it has never been defined (d'Estrée et al. 2001). Equally challenging is when evaluators are asked to show if a program achieved its results by measuring progress against a nebulous goal such as "contributing to stabilization and security." If program designers can do a good job of specifying the expected *types of change*, this offers evaluators something that they can work with to deliver insights to the program. Clarifying the type of change a program

seeks to achieve not only aids evaluators in their ability to contribute to peacebuilding, but it is also central to another key shift—the adoption of *theories of change*.

Theories of Change

Theories of change as a key evaluation concept have slowly been adopted into the peacebuilding discourse over the past twenty years. One can now find theories of change distributed throughout peacebuilding documentation, from project fact sheets to government policy. For instance, there are several references to theories of change in "Gender Equality: A Foundation for Peace—Canada's National Action Plan 2017–2022," including qualifying for mention in its Executive Summary.[11] While now a relatively common, though not yet ubiquitous, concept, this was not the case in the late 1990s.

It was not until 2001–2002 when the idea of theories of change first appeared as a topic at peacebuilding events, though one that was still vague and not entirely understood. As explained in a 2003 INCORE report on conflict resolution and evaluation, "The consensus around the importance of exploring theories of change coupled with the nascent nature of the conceptual discussion compelled the authors to look beyond the conference dialogue. The resulting research found a complete dearth of literature addressing theories of change in conflict resolution" (Church and Shouldice 2003, 30).

Since the early 2000s the peacebuilding literature has generally coalesced around the core idea of a theory of change: "A theory of how and why an initiative works" (Weiss 1995). Beyond this definition, however, there is considerable variance in what agencies think constitutes a useful theory of change. A review of practice suggests that there are three interpretations of what a theory of change is, distinguished by varying degrees of detail: *a defining strategy, a summary of core change processes*, or *a detailed articulation of all cause-effect relationships*.[12]

The most macro-level camp takes the theory of change as a defining strategy. In this interpretation the theory of change explains how a conflict will be resolved or peace maintained. For example, if we withdraw the resources of war, then the conflict will be resolved, or "peace come[s] through the transformative change of a critical mass of individuals, their consciousness, attitudes, behaviours and skills" (Church and Rogers 2006, 14). Some of the earliest work on theories of change specific to peacebuilding put forth this understanding, such as the original work of Peter Woodrow, though the adoption appears to have since waned (Church and Rogers 2006).

At the mezzo-level camp, theories of change are interpreted as a description of the core change processes that a program is based upon. In the best-case scenario, these descriptions are an executive summary of a detailed design process, as in the example below.

> If people from within the Criminal Justice Sector (CJS) who act with integrity can establish strong relationships with each other, then they will feel added protection and [be] empowered to act against corruption more openly and often, because they will have support (e.g. emotional, hierarchical, tactical) from those inside the system. *And*: If those with integrity show that resisting corruption is possible, this will encourage resistance by others who have been participating in corruption but feel it is not right, because they will know resistance is possible and they will not be isolated for doing so. *And*: If those inside the CJS are connected with islands of integrity working in criminal justice but not employed by the government, then these relationships will provide additional motivation, information and protection, because they are not under the same hierarchy as those working within. (Scharbatke-Church, Barnard-Webster, and Woodrow 2017, 5)

In the worst case, theories of change are the product of the fundraising department's interpretation of a project. When that occurs, the value of thinking through how change happens is not gained, as the framing is done for marketing purposes. While not based on a scientific review of the field, it is the author's experience that this mezzo-level camp is gaining the most traction in the peacebuilding field.

Finally, a detailed articulation of the change processes is the third and most micro-level camp. This camp articulates a detailed breakdown of every cause-effect connection within an intervention.[13] It draws from the belief that "every action we take, from the overall goal of the project to each single activity, has a theory of change behind it."[14] This means that even a simple conflict resolution training that seeks to change attitudes, knowledge, and behavior would have numerous embedded theories of change. Program planners would lay out the theory behind each activity that is intended to catalyze a shift.

It is this final detailed interpretation of theory of change that is most useful for evaluation. Lisa Schirch, a thought leader in the peacebuilding and evaluation space, stresses that explicitly articulated, well-constructed theories of change that document their adaptations as programs evolve are the cornerstones of useful M&E (2015).

One of the reasons behind the differing interpretations of what constitutes a theory of change could be the subtle but important variance one finds in guidance documents for developing a peacebuilding theory of change. The most typical guidance recommends the use of an "if-then" statement to develop a theory of change.[15] (***If*** we do this action, ***then*** this change will result.) CARE International's guidance includes a few examples that exemplify this structure: "If there is evidence of peaceful resolution of conflict, then confidence in and acceptance of peacebuilding mechanism is increased."[16] "If women occupy leadership positions inside given structures and institutions, then they will be able to bring other women into peacebuilding activities."[17]

Schirch has added a peacebuilding-centric twist to the classic "if-then" statement by making it "if-then-impact."[18] Here the "if" grounds the hypothesis in the conflict assessment rather than the activity statement. For instance, "***If*** a conflict assessment finds that government corruption is driving inequality and division among classes, ***then*** planning to support a community's capacities for monitoring and addressing corruption can ***impact*** peace and security by reducing corruption and empowering communities to participate in governance" (Schirch 2015, 2). This version

has not received significant attention, possibly due to the paucity of conflict analysis in the field.

Finally, Scharbatke-Church and Rogers (forthcoming) put forth a slightly different form: "if-then-because." In this version, ***if*** this action, ***then*** that change, ***because*** of this reason as evidenced by or due to this contextual grounded assumption (forthcoming). This version more explicitly incorporates the assumptions or evidence underpinning the change pathway without which the potential for learning through evaluation becomes minimized. When the "because" is included, one is able to test why the theory of change has or has not worked.

CHANGES IN PEACEBUILDING EVALUATION PRACTICE

These two decades saw the field making significant adjustments to its practice. From the amount of evaluation conducted to the type of evaluation practices implemented, twenty years has made a big difference.

Explosion of Evaluation Practice

The most glaringly obvious change in evaluation in the peacebuilding field in the past two decades is the explosion of evaluation activity. Using the broadest definition of evaluation, the peacebuilding field has moved from virtually no evaluation practice in the mid- to late 1990s to enough work being done that there was sufficient evidence for the International Initiative for Impact Evaluation (3ie) to produce a Peacebuilding Evidence Gap map in 2014.[19] The map provides "easy access to the best available evidence on the outcomes of peacebuilding interventions in conflict-affected settings."[20] With this eruption of activity has come many adaptations to the way in which evaluation practice is done in peacebuilding.

Proliferation of Guidance

The number of technical evaluation guidance documents for peacebuilding and programming in fragile state environments has grown exponentially in the past two decades (Kwano-Chiu 2011). Consider the fact that the very first writing on the issue was produced in 1997/1998, and now the Design, Monitoring and Evaluation for Peace portal (see Oatley, this volume) offers 1,123 resources (as of 2018).[21] This growth is not only apparent in the array of written materials but also apparent online and in in-person training and professional engagement opportunities.

One of the most widely influential technical guidance documents has been the Organisation for Economic Co-operation and Development's Development Assistance Committee's (OECD-DAC) *Evaluating Peacebuilding Activities in Settings of Conflict and Fragility Guidance Note*. Published in draft in 2008 and finalized in 2012, this document "guide[s] policy makers, field and desk officers, and country partners towards a better understanding of the role and utility of evaluations."[22] While the guidance note covers a range of important issues, such as the need to conduct a conflict analysis and articulate theories of change, one of its critical contributions to the field is the translation of the standard OECD-DAC evaluation criteria[23] to the peacebuilding realm. This provides peacebuilding-specific interpretations of what should be reviewed in an evaluation.

Variety of Evaluation Processes

Not only has more technical content become available, peacebuilding evaluators are now far more likely to draw from the thinking in the broader evaluation discipline. As of ten years ago, there was little-to-no interaction between the professional evaluation community and the evaluators in the peacebuilding arena, but that is changing.[24] One of the results of the interaction is the greater variety of evaluative processes that have been adopted by the peacebuilding community, though one could not argue that

alternative approaches are common practice or even well integrated into M&E systems yet.

Three evaluative processes—evaluability assessment (EA), meta-evaluation, and synthesis—offer different benefits than formal program evaluations and have started to be utilized by peacebuilders.

1. *Evaluability Assessment*: An EA looks at the clarity of the program design, the availability of data, and the context of the program to determine if the program is conducive to evaluation. In effect, it is an analysis of the possible return on investment of doing an evaluation—how much will the agency learn or gain—against the perceived costs and difficulties.[25] While peacebuilding agencies have only started to dabble with EAs, the process received a considerable boost in attention after the Department for International Development (DFID) released a 2013 review articulating the benefits of the approach (Davies 2013).
2. *Meta-evaluations*: Meta-evaluations assess the quality of the evaluation process(es) and product(s). They are typically initiated to identify ways to improve the quality of evaluations in an agency or within a type peacebuilding. For instance, in 2015, the Peacebuilding Evaluation Consortium conducted a review of inter-religious peacebuilding program evaluations in order to "understand what the current trends are in the evaluation of inter-religious peacebuilding programs and to assess the quality of evaluations" (Vader 2015, 3).
3. *Synthesis*: Synthesis reviews the evaluations of a particular type of peacebuilding (e.g., violence reduction or peace journalism) to accumulate findings and identify common lessons and/or results applicable to the programming. While a meta-evaluation focuses on the evaluation quality, a synthesis seeks to compile lessons from the evaluative conclusions. This process goes beyond the project-centric nature of a single evaluation to draw lessons more relevant to the wider field. For instance, in 2013, the Conflict Management and Mitigation Office at USAID initiated a meta-evaluation and synthesis of its People-to-People reconciliation programming in fragile states, in

an effort to understand what difference these programs were making collectively and the quality of the evidence underpinning those conclusions.

The settings for peacebuilding evaluations are highly dynamic, complex, and influenced by interdependent factors. While the peacebuilding evaluation field has benefited from greater engagement with the evaluation discipline, these unique challenges of fragile and conflict-affected states also require evaluation approaches—the theories that act as conceptual underpinnings to the evaluation—to be fit for purpose.

Approaches like *Outcome Mapping*[26] pioneered by Canada's International Development Research Centre and *Developmental Evaluation* formalized by Michael Quinn Patton (Patton 2010) draw from systems and complexity thinking. *Outcome Harvesting*, developed by Ricardo Wilson-Grau, rejects the prescriptive notion of conventional evaluation and focuses solely on proving the outcomes the implementing team claims to have achieved (Wilson-Grau and Britt 2013). Finally, *Most Significant Change* (courtesy of Rick Davies) skirts the controversies associated with indicators and offers a participatory means to understand what has mattered most to a community (Davies and Dart 2005). While these approaches were not developed specifically for peacebuilding, they have been guided by the need to deal with many of the challenges that peacebuilding faces, such as complexity, power differentials, and high uncertainty.

Technology and Data

The most recent trend to impact the evaluation of peacebuilding field is the integration of a vast array of technologies and the abundance of resulting data. Predominantly utilized in the data collection stage of an evaluation, technology has enabled evaluations to significantly increase the amount, quality, reach, and real-time nature of data collection without comparable cost increases. Internet-connected tablets enable real-time data quality review and eliminate the need for a second data entry stage. Social media

offers a way to connect with the hard-to-reach, and drones allow images to be taken from places too dangerous for evaluation teams to travel. For instance, a drone could be used to determine if poppies were still being cultivated by rural farmers in Afghanistan, even if it was too dangerous for an evaluation team to travel to the site.

The digital reality of data collection is forcing agencies to think more carefully about the security of their online systems, and by extension, their data. Possible security breaches are no longer a hypothetical threat. In 2017, for instance, Catholic Relief Services (CRS) discovered the system they were using for "digital payments in West Africa had been hacked, exposing personal, geographic and photographic data about its beneficiaries" (Abrahams 2018). While this breach involved programming data, the evaluation relevance should be clear. If an evaluation team is collecting confidential information from key informants and uploading that data to a digital system hosted by those being evaluated, what measures must the evaluation team take to ensure that the confidentiality of the data is honored?

In response to these new challenges, the past few years have seen the larger multimandate agencies develop data policies to lay out clear expectations and processes regarding appropriate uses of data and transparency.[27] This is a bigger challenge for smaller organizations without the same internal expertise. While policies are an important start, according to Karl Lowe, chief information officer at CRS, "the biggest challenge is not necessarily the technology—it's the people and processes associated with the technology, and then driving a culture of using data in an appropriate way" (Abrahams 2018).

The fragile contexts of most peacebuilding programs create some unique considerations when it comes to capitalizing on the benefits of technology, while minimizing the risks. Concrete factors tend to play more of a role in fragile environments compared to development contexts, such as whether or not local script is included in the application or software, or accessibility of electricity and reliability of the Internet. There are also greater risks in fragile environments for government action to inhibit a tech-enabled

evaluation. Governments that are party to internal conflict are fully aware that controlling information flows is in their interest. For instance, in the first month of 2019, five African governments temporarily shut down Internet access amid political unrest. Anyone implementing an Internet-based data collection process at that time would have been stalled until the government stopped its action, unless they had robust contingency plans in place.

Ensuring that technology does no harm is an oft-forgotten element of a tech-enabled evaluation (Scharbatke-Church and Patel 2016). What are the personal safety implications of handing out smartphones to villagers to collect data or take pictures? What could happen if data stored on the cloud were accessed by a party to the conflict? What does an evaluation team do in the case of theft of tablets by locally sourced enumerators? These, among many other considerations, need to be taken into account when integrating technology into a peacebuilding evaluation.

New Actors in the Evaluation Arena

The new technology (hardware and software), security challenges, and explosion of digital data have introduced a number of actors into the evaluation conversation that would not have participated even as recently as 2010. Significant among these new players is the increased importance of information technology (IT) departments. M&E professionals now must routinely engage IT professionals as they develop internal systems or perform an external evaluation function. They are not the only new actor in the M&E space. Big data, for instance, require mathematical and software expertise that goes far beyond a typical M&E professional's competency. This brings the M&E field into contact with different actors, such as CivTech activists or data analytics, groups who were not on the invitation list a decade ago.[28] This requires peacebuilding evaluators to learn how to communicate and interact so that evaluation is optimized by technology.

While many of the changes enabled by technology and the data explosion over the past five years have been beneficial, the issue for the peacebuilding

evaluation field is to ensure that the technology enables evaluation rather than drives it. This is particularly true from a conflict-sensitivity perspective, to ensure that an evaluation does not exacerbate conflict dynamics. It is important that the evaluation purpose and audience continue to drive the design and implementation of evaluations and not the latest technological gadget. For instance, when an evaluator starts with the desire to use the latest cool social media platform as a means to collect data, rather than focusing on the evaluation question they need to answer, the cart has been put before the horse.

Internal Expertise

Turning our attention to changes in agencies themselves, the focus on results and the requisite skills and processes that this focus requires of agencies has meant that peacebuilding implementers have had to bolster their internal evaluation expertise. This has led to the creation of positions and internal M&E units in peacebuilding agencies. Not surprisingly, the larger, multi-mandate agencies that include peacebuilding were first past the post, with CARE establishing an M&E department in 1995 and CRS creating its first M&E position in 1998. The UN Department of Peace Operations (formerly DPKO) was part of the vanguard, opening its Lessons Learned Unit in 1995. The smaller single-mandate peacebuilding agencies created these roles a bit later. For instance, Search for Common Ground, one of the largest INGOs with a peacebuilding mandate, was one of the first to formally institute an evaluation position in 2003.

WHAT HAS NOT CHANGED ENOUGH

In many ways, the field of peacebuilding evaluation has grown unrecognizable in the span of twenty years. From a new results paradigm to the adoption of a wide range of evaluative processes, the changes are remarkable in many respects. Yet this has not been a wholesale

transformation, as there are a number of aspects that have stayed consistent. The remaining of this chapter will discuss the continuing gaps in practice.

Evaluation Purpose versus Practice

Over the past twenty years, the ability of peacebuilding evaluation to deliver on its purpose has not seen the same rate of improvement as evident elsewhere in the international development and humanitarian fields. While clear improvements have been made, peacebuilding evaluations do not yet deliver, as a consistent norm, performance accountability or learning to its primary audience—which the evaluation field writ large agrees is evaluation's *raison d'etre*.[29]

Performance accountability in its simplest form probes whether the program or agency made the difference it set out to do. A performance accountability interpretation asks: what did the program *change in the conflict*? In this sense, change refers to making a difference on the ground (i.e., outcomes or impact) (Ebrahim 2010). This is not to be confused with compliance, which assesses whether the program accomplished the work (e.g., activities and outputs) originally proposed. Compliance does contribute to performance accountability, but they are not the same. The latter focuses on change on the ground, as clearly differentiated from the effort that was expended.

Learning, from the evaluation perspective, is the acquisition of knowledge or skills that could be used to improve programming. It is commonly conceptualized according to the following categories:

- Single loop: learning about tactics within the intervention. This looks at whether the program was able to accomplish its stated goals; in other words "doing things right."
- Double loop: learning about assumptions and strategy behind a program. This explores whether the program was the right one at the right time with the right approach or "doing the right thing."

- Triple loop: learning about how we learn. This examines how peacebuilders decide if they are doing the right thing or if they need to change their approach.[30]

In the late 1990s, the fields of evaluation and peacebuilding (then called international conflict resolution) were just meeting, and so expecting these processes to deliver quality performance accountability and learning results would have been a very high bar to set. The rarity of peacebuilding evaluations happening at this time also meant that there was no significant terrain for testing and improving the evaluation processes themselves. The situation was made even more difficult by the prevailing attitude at the time, which was highly resistant to the evaluation enterprise.[31] This meant that sufficient resources—time and money—were rarely given to evaluation processes that would have enabled them to deliver on the purpose.

The past two decades have seen large advancements in evaluation practice resulting in vast improvements in the technical sophistication of evaluations, but without commensurate improvement in contributions toward performance accountability and learning. The possible exception to this may be single-loop learning resulting in improvement in tactics within programming, which has come about partly due to the degree of attention given to M&E tools over the past twenty years. While single-loop learning is necessary and useful to practitioners, it is not sufficient to fulfill the learning purpose or the potential of evaluation.

The *disconnect between the evaluation practice advancements and the purpose achievement* has multiple causes. Three of the more significant factors will be explored here, recognizing that each plays out differently within peacebuilding agencies themselves.

The first is the most obvious, whereby what should be driving evaluation—performance accountability and learning—is replaced with donor compliance requirements and/or the needs of fundraising and communication directors. When ascertaining donor compliance (i.e., did the agency do what it said it would do) becomes the *de facto* evaluation purpose, evaluations can end up focusing on proving that activities

happened and outputs were produced, rather than on contributing to understanding whether and how change happened (Natsios 2010). When fundraising departments request that evaluations include stories (to put a human face on the peacebuilding work, which can be featured on the website), this can drive the data collection methodology decision toward thin case studies purposefully selected to offer positive frames. While these demands are real, when they become the purpose of an evaluation—explicitly or implicitly—this distorts key evaluation decisions and processes, reducing the evaluation's ability to serve performance accountability and/or learning.

Second, for learning and performance evaluation to be fully embraced by evaluation processes, these concepts must be present in the fabric of the peacebuilding organization itself. In other words, if one expects a project evaluation to authentically engage in an assessment of what difference the project made (i.e., change in the conflict) or to catalyze learning, then the organization itself must signal that these things matter, through their actions, priorities, and internal processes. These concepts, often referred to as promoting a "culture of evaluation" or becoming a "learning organisation," go well beyond individual project evaluation.[32]

As Preskill and Torres, two thought leaders in the discipline of evaluation, lay out in their "Readiness for Organizational Learning and Evaluation Instrument (ROLE)," agencies need to look internally at six categories to determine if they are ready to engage in evaluation. These include an agency's culture (collaboration, risk-taking, participatory decision-making), leadership, systems and structures, communication, teams, and evaluation processes and practices (Preskill and Torres 2016). Readiness to engage implies that an organization has the internal culture, processes, and skills to benefit from the evaluation itself.

While there are clearly exceptions, typical peacebuilding organizations do not have a culture of evaluation. Data do not have equal weight with experience in decision-making; curiosity and inquiry into why things happened is not prioritized; and change on the ground receives second-order billing to fundraising accomplishments or media references.[33] As

Swisspeace states, "To make results-orientation a worthwhile effort, not only the tools, but also the institutional setting in which they are applied needs to be rethought. Processes of institutional learning need more attention" (Bachtold, Dittli and Servaes 2013, 14).

The third reason that evaluation often fails to meet performance accountability and learning expectations lies in the insufficient discourse over the past two decades in the peacebuilding community about *accountability*.[34] It is difficult for evaluation to be an effective tool of accountability when the field lacks clarity on what constitutes accountability: what it means, with whom the accountability relationship(s) exists, and how an agency would use evaluation to satisfy it.

There is a tremendous amount of literature discussing accountability from a variety of perspectives: private sector, civil society, and government.[35] While there is no single generally accepted definition, *quality*, *obligation*, *willingness*, *and responsibility* are all qualities typically associated with the concept. Unfortunately, a punitive connotation is often associated with the term, although that is more a result of how accountability is implemented rather than the meaning of the concept itself. Boiled down to its essence, accountability can be understood to mean *a clear commitment that—in the eyes of others—has been met*.

Using this framing, the first challenge that arises is with establishing what commitment has been made and to whom. Starting with *upward* accountability— that is, between a peacebuilding agency and its donor—the issue of the donor's preference for compliance around those things that can be observed or counted becomes the challenge. This overly narrow compliance interpretation is clearly seen in the former USAID administrator Andrew Natsios's widely cited piece on "the Counter-Bureaucracy" (2010). Here he states, "Accountability should not be confused with developmental effectiveness. A program can be highly accountable with no fraud or abuse, yet be a developmental failure; conversely, a program can suffer from a leakage of funds and poor recordkeeping and yet be highly innovative and successful developmentally" (Natsios 2010, 37). This statement suggests that the "clear commitment" is more about budgetary compliance than

delivering results. While problematic, this challenge with the meaning of accountability is not limited to the peacebuilding community but plagues the entire development industry.

Where the peacebuilding community experiences more unique challenges is when *downward* accountability is introduced, where the "eyes of others" refer to the people living in/affected by the conflict. Kosovo in the early 2000s offers a clear example of the challenge of downward accountability to peacebuilding (CDA 2016). In that postwar setting, the donor community coalesced around a common policy of a multiethnic Kosovo and thus supported "minority return." Due to the demographics of expulsion, this predominately meant the return of the Serb population to areas that had become under the control of the Kosovar Albanian "state." Donor support for peacebuilding was dedicated almost exclusively to multiethnic projects, such as multiethnic agricultural cooperative and/or facilitating returns. Yet this was not the vision of Kosovo that a majority of either the Serb or the Kosovar Albanian population had in mind. How does evaluation assess performance accountability from a downward accountability perspective, when "the clear commitment" is not one wanted by the people it claims to serve? This challenge explains why the humanitarian sector, in the same time period, engaged in a field-wide accountability dialogue that resulted in a significant sectoral initiative—while the peacebuilding community has effectively avoided engaging in the thorny issues that the accountability question raises for the field.[36]

Inconsistent delivery on performance accountability and learning feeds a vicious cycle that reinforces the challenges peacebuilding evaluations have to deliver on their purpose. When evaluations fail to provide practitioners and their organizations with programmatically useful insights, assumption challenges, or evidence-backed conclusions, this reinforces a perception that evaluation costs are not commensurate with the benefits. This puts downward pressure on evaluation budgets, as money is viewed to be better spent doing "more" peacebuilding activities, which then keeps evaluation under-resourced. Poor evaluations are rarely able to design proper methodologies, engage in adequate user education, or develop participatory

feedback processes; instead, they typically feed the donor machine or generate pithy website quotes. This reinforces the perception that these kinds of results are what evaluations are best meant to do, which in turn diminishes any incentive to adopt a culture of evaluation. This reinforces the sense that evaluation is not a priority process central to good peacebuilding practice.

Indicators

Twenty years ago, indicators for project-level results were not well understood by peacebuilding implementers, and therefore commonly not included in the M&E plans. Donors, for their part, were embarking on the first of what would become several rounds of infatuation with universally applicable indicators.

Indicators are proxies; they signal a change that cannot be observed or measured directly (Church and Rogers 2006). The passing of a law guaranteeing equal rights to all citizens, for instance, does not require an indicator to prove that it exists. The law is passed or not. The degree of trust in another group, on the other hand, is more difficult to measure directly, so a program may opt to use indicators to gain a sense of whether or not this change is happening. Indicators should be generated during program design, and use data collected (and verified) as part of monitoring or originally collected for an evaluation.

Today, a review of proposals from three different peacebuilding organizations would likely reveal three very different interpretations of indicators.[37] One proposal might use indicators as signals of change (i.e., a proxy), while a second might consider indicators as synonymous with activities achieved (e.g., fifteen peace huts built) or as output statements (e.g., fifty people in attendance at conflict resolution training). The third proposal might use indicators as markers of progress toward an end result, for instance, "we will build five of the 15 peace huts by the end of the 6 months." Depending on the donor template that was used to format the M&E plan, it is possible to find all three of these approaches intermixed.

This wide range in interpretation of what an indicator is remains as true today as it was in the late 1990s.

Using the logical framework (i.e., logframe) as a program design tool has contributed significantly to this confusion, due to its matrix structure. The logical framework was initially introduced by donors as a way to get consistent, synthesized proposals. Implementing actors have since attempted to use the logframe as a program design tool. The matrix structure applies the indicator column across all rows, with the rows dedicated to activities, outputs, and change statements (e.g., goal and objective). If one attempts to fill in the indicator column consistently across all of these levels—all rows in the matrix—then the marker of progress interpretation is the only one possible. Yet the purpose of indicators is to act as a proxy for a nonmeasurable result.

Since we lack common agreement about what an indicator is, it becomes difficult to generate ones that are accurate. The accuracy challenges are exacerbated further by the fact that most indicators are developed in INGO headquarters (HQ) in the Global North, and therefore do not reflect the culture and context in which they are to be applied (Mac Ginty 2013b; Firchow and Tilton, this volume).

This eclectic assortment of HQ-generated indicators found in the M&E sections of proposals leave those responsible for implementing M&E frameworks in a difficult position. They can (1) gather the information promised, knowing that it is either inaccurate and/or not useful; (2) expend significant time and political capital in an attempt to develop indicators that are accurate; or, finally, (3) refuse to use them at all. The latter option may explain why, in the author's opinion, one finds considerable presentation of indicators in proposals—yet far less actual indicator data in monitoring or evaluation reports.[38] The absence of indicator use in peacebuilding evaluations could also be due to a lack of commitment to gather data, the "compliance as accountability" mindset, poor budgeting, lack of skill, or inexperienced evaluators. Here again, at the trend level, the degree of indicator use (or lack thereof) appears to be relatively consistent across the past twenty years.

Despite this, the enthusiasm for indicators remains strong, and nowhere more so than within the donor community. Signatories to the New Deal for Fragile States, for instance, committed to develop a set of indicators for each goal to track progress at both global and country levels. Donors have been particularly fond of universally applicable indicators—those that can be used across contexts—despite a general consensus that indicators need to be contextually derived.[39] The UK DFID was a leader in this regard in the late 1990s,[40] with the U.S. State Department adopting the concept with zeal under the leadership of Secretary of State Condoleezza Rice and the development of the standard foreign assistance indicators (aka the "F-indicators").[41] The F-indicators applied to the "Peace and Security" program area illustrate the lack of consistency used in indicator interpretation. They offer both "number of USG-funded events to build support for peace" and "percent of political districts/provinces that are not a major source of narcotics cultivation" as potential indicators. The first represents an activity, while the second represents a change.

Though led by the bilateral donor community, work on universal indicators is also underway in NGOs. For instance, in 2010 CRS had produced the globally accepted indicators for peacebuilding, which offers eighteen illustrative indicators felt to be appropriate and effective for project M&E.[42] For instance, one indicator posits "the number of joint initiatives between youth organisations and strategic governance agencies in x time period," while another suggests "primary school net enrolment for girls," to name just two. While framed as "globally-accepted," CRS stresses that there is "little emphasis on the use of the indicator outside of the specific project or program context."[43] Instead they offer a template from which a practitioner could work to generate their own indicators, inspired, in theory, by the example indicators provided. This attempts to bridge the inherent tension found between the desire for universal indicators and the generally accepted principle that indicators must be context-specific.

CONCLUSION

After twenty years, the field of peacebuilding evaluation has undergone tremendous change. The explosion of engagement, partially caused by the results orientation adopted through the international community's Aid Effectiveness agenda, has caused almost every peacebuilding organization to create an in-house position dedicated to evaluation, complete with a related policy suite, such as an evaluation policy, data policy, and so forth. Technical guidance, offering step-by-step instruction, exists on everything from crafting theories of change to writing evaluation terms of reference. Meanwhile, the amount of data generated has increased exponentially, and technologically enabled data collection within an evaluation is the new normal.

Tremendous change has occurred, while aspects of the field lag behind. Despite adoption of "results" as the new norm, there is no commonly accepted definition of results. The donor community has tended toward systematized processes that prioritize tangible products (i.e., outputs), which has reduced risk tolerance and innovation in peacebuilding. An alternative understanding of results consists of *change in the conflict context*, separate from the work done by the peacebuilding agency. Differentiation among types of change is now part of the peacebuilding lexicon, which is a clear difference from twenty years ago, although this is not yet the commonly accepted meaning of results.

A broader range of evaluative processes are now being conducted by peacebuilding agencies, and some strides have been made in the diversity of evaluation approaches applied. These are all important improvements, that center on the tactical application of a program evaluation. The field has not seen a commensurate change in the ability of evaluations to deliver on performance accountability and learning, the driving forces behind evaluation. Fundraising, public relations, and donor compliance are far too often the main drivers behind evaluation. The peacebuilding field's discomfort with an authentic conversation around accountability continues to limit the ability of evaluative processes to deliver genuine accountability or double/triple-loop learning.

Looking forward to the next two decades, there are many exciting opportunities and daunting challenges facing the field of peacebuilding evaluation. Opportunities will emerge as the field continues to become more sophisticated in its technical application. Peacebuilding agencies may choose to distinguish themselves through the adoption of more robust evaluative cultures, as a competitive advantage for accessing resources. There is also real potential for synergies to arise between evaluation and peacebuilding programming objectives through the adoption of greater indigenous methodologies. Evaluators could also lead a conversation around what it means to decolonialize peacebuilding by asking the simple question: who decides what is success?

Conversely, the technologically driven explosion in data and its resulting concerns regarding consent, confidentiality, and privacy will continue to pose challenges for evaluation. While over-access to information will exist, in other cases the shrinking civil society space will result in restricted access. As governments in many fragile states increase the barriers to conduct research through country-led research ethics processes and affiliated research-specific visa requirements, their ability to politically influence where and what gets evaluated will increase.

As long as there is conflict, there will be a need for peacebuilding that works. Given the trends of today, there is little suggestion that this maxim will change in the next two decades. The need for peacebuilding evaluation will continue, as will its evolution.

NOTES

1. Council on Foreign Relations (n.d.).
2. Institute for Economics and Peace (2018), 26.
3. World Bank Group (2018).
4. Practitioners' Network for European Development Cooperation (2017).
5. Ibid.
6. Fourth High-Level Forum on Aid Effectiveness (2011), para. 3.

7. United States Agency for International Development.

8. The First High Level Forum on Aid Effectiveness was held in February 2003 in Rome. It brought together heads of multilateral and bilateral development institutions along with aid recipient countries in an effort to harmonize their policies and practices.

9. The Istanbul Principles are the foundation for the International Framework for CSO Development Effectiveness, adopted in June 2011 at the Second Global Assembly in Cambodia. The frameworks sets out guidance for interpreting and aligning CSO practices with the Istanbul Principles in diverse local and sectoral settings. For a complete account of how the principles and framework came to be, see Tomlinson (2012).

10. New Deal (n.d.).

11. It is acknowledged that the presence of theories of change in documentation is not synonymous with saturation at the field level where frontline staff work. Of interest is the Care International UK and International Alert "Strengthening Capacity to Design, Monitor and Evaluate Peacebuilding Programming" project, where they found that most field staff were not familiar with the concept in 2010–2012.

12. Woodrow puts forth an alternative analysis, positing that there are three levels of theories of change that can work together. See Woodrow with Oatley (2013).

13. See, for instance, the Care International UK and International Alert "Strengthening Capacity to Design, Monitor and Evaluate Peacebuilding Programming" project.

14. CARE International UK (2012), p. 3.

15. See, for instance, CARE International UK (2012) or Stein and Valters (2012).

16. CARE International UK (2012), p. 3.

17. Ibid., p. 9.

18. This aligns to some of the theory of change writing from the evaluation field, which posits that the "if" should ground the theory in an analysis of the problem.

19. While this is an achievement of significance, it needs to be recognized that there are a large number of areas of the map without evidence and that the evidentiary work falls on the more research-heavy side of the research-to-evaluation spectrum. Nonetheless, the mere existence of the map would not have been possible as recently as 2010.

20. International Initiative for Impact Evaluation (3ie) (n.d.).

21. Some of the first work can be traced back to Marc Howard Ross and Jay Rothman, with the development of Action Evaluation around 1997. ALNAP also contributed an early seminal piece (Spencer 1998, cited previously in this piece). See Search for Common Ground's "Learning Portal for Design, Monitoring and Evaluation for Peacebuilding" (http://www.dmeforpeace.org/), which shares resources and learning events and convenes a community of practitioners committed to the sharing of best and emerging practices in DME for Peace.

22. OECD iLibrary, DAC Guidelines and Reference Series (2012). (This quote is from the website: https://www.oecd-ilibrary.org/development/evaluating-donor-engagement-in-situations-of-conflict-and-fragility_9789264106802-en.)

23. The primary evaluation criteria put forth by the DAC for all fields include relevance, sustainability, effectiveness, efficiency and impact. The definitions, however, are tailored to the specific area, for example, humanitarian or development.

24. By using the American Evaluation Association (AEA) annual conference searchable program as a proxy to assess the degree of integration of peacebuilding evaluation with the broader discipline, the slow cross-fertilization of peacebuilding evaluation with the wider field is evident. In 1999, there was one related entry, a poster presentation, titled "Conflict Resolution: A New Frontier in Evaluation" by Jennifer Holt of the University of Minnesota. By 2017 there were five complete panels, which would constitute between fifteen and twenty-five individual presentations.

25. U.S. Department of State—Bureau of Conflict and Stabilization Operations (2019).

26. For more information on Outcome Mapping, join the vibrant online "Outcome Mapping Learning Community" found at https://www.outcomemapping.ca/.

27. In April 2018, 3ie produced their initial 3ie Research Transparency Policy (http://www.3ieimpact.org/sites/default/files/3ie-policy-research-transparency-policy.pdf) which details the actions expected of researchers when it comes to transparency. The 3ie is not alone; Oxfam developed a Responsible Program Data Policy, in 2015, that lays out their expectations on the responsible use of data (https://policy-practice.oxfam.org.uk/publications/oxfam-responsible-program-data-policy-575950). In terms of data protection, the ICRC released, in 2017, the *Handbook on Data Protection in Humanitarian Action* (https://www.icrc.org/en/publication/handbook-data-protection-humanitarian-action). It seeks to help humanitarian organizations comply with personal

data protection standards, by raising awareness and providing specific guidance on the interpretation of data protection principles in the context of humanitarian action, particularly when new technologies are employed.

28. MERL Tech (http://merltech.org/) is one of the conveners of these two fields—developing resources, hosting events, and promoting effective technology integration. Build Peace (https://howtobuildpeace.org/) offers an annual event linking practitioners, activists, academics, policymakers, artists, and technologists from around the world to share experience and ideas on innovations in peacebuilding.

29. This assertion is drawn from two primary experiences. First, the author's experience as a M&E advisor to the peacebuilding work of two significant bilateral donors. In this role, the author has been privy to a large cross-section of evaluations. Second, students in the author's Advanced Evaluation course at the Fletcher School of Law and Diplomacy conduct ten to fifteen meta-evaluations annually.

30. Single- and double-loop learning concepts were originally developed by Chris Argyris and Donald Schon. See also d'Estree this volume.

31. This resistance did not change overnight in the peacebuilding field. In the early 2000s, Martin Honeywell, the deputy secretary general of International Alert from 1997 to 2003, wrote a paper describing the resistance of the peacebuilding community as changing from overt resistance to *resistance that there was even resistance* to evaluation.

32. See, for instance, Argyris and Schon (1978); Levitt and March (1998), 320; Britton (1998).

33. This conclusion is based on the author's experience in developing and implementing the Besa Effectiveness Assessment Tool for High-Performance Social Change Organizations (https://www.dmeforpeace.org/wp-content/uploads/2017/06/214_Besa20Working20Paper20120Effectiveness20Assessment20Tool20Overview.pdf), which assesses the degree to which an agency views change on the ground as mission critical. We worked with this tool across a number of donors and implementers in the peacebuilding field over the span of ten years. It draws heavily from the Preskill and Torres work as well as evaluation capacity-building and change management.

34. In writing this chapter, the author did a significant literature search to identify work that focused on accountability in peacebuilding, and was able to identify one dedicated piece of work, a chapter in Bush and Duggan (2015).

35. See Wenar (2006), 1–23; Winters (2010), 218–243; "What Is Accountability, Really?" in Scharbatke-Church and Rogers. 2019 (*forthcoming*), available at https://www.walkthetalk.com/media/sneak_a_peek/lookinside/winning_with_accountability_01.pdf.

36. The Sphere standards, initiated in 1997, were one of the earliest initiatives to improve the quality and accountability of the humanitarian sector. According to its website, the Sphere standards are the most widely recognized humanitarian standards across the globe (https://www.spherestandards.org/humanitarian-standards/).

37. This assertion is drawn from a number of professional engagements that the author has undertaken over the past decade that have given her access to a wide number of peacebuilding proposals. This includes, but is not limited to, acting as a reviewer for proposals submitted to bilateral and multilateral donor institutions and evaluating programs. It also includes work done to assess the culture of evaluation within major peacebuilding organizations, whereby program designs (i.e., proposals) are one source of insight into the level of evaluative thinking expected in the organization.

38. This statement is based on the review of ten to fifteen meta-evaluations annually over the past decade, involving 80–150 evaluation reports.

39. Catholic Relief Services (CRS) (2010).

40. DFID's attempt did not go far; to the author's knowledge, no universal indicators are used for peacebuilding projects today. That said, as recently as 2015, DFID commissioned a study on indicators for conflict, stability, security, justice, and peacebuilding.

41. For the full listing of F-indicators, see https://www.state.gov/f/indicators/.

42. CRS (2010).

43. "Introduction," CRS (2010), 5.