# *Chapter 1*

# A Refresher on Evaluation

Tamra Pearson d'Estrée

They say, "A little knowledge is a dangerous thing." Evaluation seems to be one of those topics on which many people know a little, and base their conclusions on that, or worse, feel they are equipped to train others. However, it is difficult to have the important conversations that peacebuilding practitioners—and supportive academics—need to have about evaluation if we start from partial (and sometimes mis-) understandings about the nature, purpose, practices, and tools of evaluation. So this chapter begins with a call to begin from the same page and presents a primer and refresher on the basics of evaluation.

## WHAT IS EVALUATION?

Obviously, *evaluation* means to evaluate or to assign value. But evaluation is a many-splendored thing, with many uses and many forms. Most people know of the sort of evaluation that involves judging the value of something

—was this useful? Was it worth the cost? Do we recommend trying this again or even trying it elsewhere? However, evaluation—or "the making of a judgment about the amount, number, or value of something"—can be useful for other goals. Patton (2012) groups these into six primary types and includes questions for each.

## Formative Evaluation

Before a project or program should be judged on its value, it should be allowed to develop, evolve, and improve in response to feedback. *Feedback* is a fundamental principle of evaluation and focuses on evaluation's role in facilitating *learning*. Whether you are an amoeba, a thermometer, or a program manager, one cannot successfully adjust and adapt to one's environment without feedback. How can we guide our program to better achieve its desired results? What midcourse corrections are needed? Classic formative evaluation questions include: What is working and what is not working well? What is working for some and not for others? How can impacts be increased, quality be improved, and costs be reduced?

## Summative Evaluation

At some point, people will ask, "Was this worth doing?" Summative evaluation involves making a judgment about a project or program's effectiveness, merit, or worth. While formative evaluation is improvement-oriented (and internally facing), summative evaluation is judgment-oriented (and often externally facing). This central distinction in evaluation (Scriven 1967) allows for a recognition of its different goals, uses, and users. Once a program has had an opportunity to develop a model and refine it, outsiders such as funders and fellow program developers will want to know the degree to which the program has achieved "success," which will in turn depend on how success has been articulated for that approach (d'Estrée et al. 2001). Classic summative evaluation questions include: To what degree are needs being met? Does the program provide value for money, and how

does it compare to other options? Can the outcomes be attributed to the program? Should the program be disseminated and/or scaled up?

## Monitoring

All types of evaluation need data. In addition to data being supplied to answer the aforementioned questions, managers and other users will want regular updates on basic program functioning, both to determine if adjustments are needed and if things are proceeded according to guidelines and commitments. Performance indicators can thus provide data for *monitoring* and for *accountability*. Monitoring provides data for ongoing management purposes. Classic monitoring questions include: Who is participating? What are the participation rates and how do they vary? Are there variations across subgroups?

## Accountability

Separating accountability from monitoring (Patton 2012) helps to highlight once again the difference in goals, uses, and users for different types of evaluation. While data for monitoring tell managers what might need to be adjusted, data for accountability tell stakeholders that the desired participants are involved, providers are qualified, and funds are being used appropriately. Classic accountability questions include: Are targets being met and indicators showing improvement? Are resources being allocated appropriately? Are the appropriate participants involved? Are there quality control mechanisms?

## Knowledge-Generating Evaluation

The aforementioned forms of evaluation apply to specific programs. However, we often would like to know if there are lessons to be learned by looking across similar programs. Accumulating the evaluation results across programs, sometimes called *cluster evaluation*, can lead to the identification

of patterns of effectiveness—"best practices" and "lessons learned." This is a form of research with an evaluative focus that sets out to create generalizations. Questions here include: What patterns and principles can be extracted across sites and programs? Are there results that can be triangulated "with social science theory, expert opinion, practitioner wisdom, and participant feedback" (Patton 2012, 131)?

## Developmental Evaluation

One of the more recent forms of evaluation to evolve, developmental evaluation is designed for organizational and program development (Patton 2010; 2012). While formative evaluation seeks input to improve a model or intervention, developmental evaluation seeks input to continually adapt the model or intervention to changing conditions. Summative questions of goal attainment, model replicability, and clarity of causal specificity applied to perfected and "stable" models are not useful in "unstable" contexts such as those of emergent social innovation and systems change, or in highly volatile environments like that of many peacebuilding interventions. In such contexts, evaluation serves to feed the voracious appetite for feedback and learning necessary to operate in such dynamic and evolving environments. Ongoing information gathering contributes to ongoing intervention and model development. Developmental evaluation questions include: "How is the program as an intervention system connected to and affected by larger systems in its environment? What are the trends in those larger systems? What does feedback show about progress in desired directions? What can we control and not control, predict and not predict, measure and not measure, and how do we respond and adapt to what we cannot control, predict, or measure?" (Patton 2012, 130–131).

Each of these types of evaluations has different goals and, while some overlap may be possible, for the most part these approaches should be done separately. For example, collecting data for the formative goals of identifying where shortfalls are so that staff can know what needs to be improved may be difficult to collect at the same time as data demonstrating

the effectiveness and value of a program to funders. Clarity about the purpose and goals of evaluation must be a part of the planning and design process. Evaluation should be planned into a project or program from the beginning, to allow for the collection of baselines, to create a learning culture, and to design in different forms of feedback at different points in time, for many different useful purposes.

# EXPLAINING EVALUATION

## How Does Evaluation Compare with Research?

Answering this question depends on what you consider to be evaluation and what you consider to be research. There is evaluation that looks much like traditional social science research, and there is applied social science research that is used for evaluation. However, a fundamental distinction that is made by Cronbach and Suppes (1969) is that research is conclusion-oriented inquiry, while evaluation is decision-oriented inquiry. Evaluation gathers information that will inform decisions and action. Often this means that it is very context-specific. Decisions are made for a particular context, meaning that the more generalized and probabilistic findings coming from traditional social science approaches may not be as useful, or useful in a different way from evaluation findings. This difference also means that evaluation findings may be less generalizable to new situations than findings from traditional social science research, although there again, we are depending on probabilities.

Traditions for practice- or decision-oriented inquiry have ancient roots just like scientific inquiry aimed at developing general laws. Aristotle proposed three types of inquiry: theoretical (from *theoria*) seeking truth, poietical (from *poiesis*) seeking production, and practical (from *praxis*) seeking action. The Lewinian tradition of action research sought to study social systems by changing them, and sought to provide assistance to the client system (Lewin 1948). It sought to contribute both to social science knowledge and to social action. Argyris, Putnam, and Smith (1985)

working on action science consider how to create communities of inquiry within communities of practice: those engaging in practice and using feedback, reflection, and deliberation to improve practice. They highlight how this form of inquiry is both explicitly normative, and is in the critical tradition, that is, "a critical social science engages human agents in self-reflection in order to change the world" (6).

Both qualitative and quantitative methods can be useful for evaluation, in the same way that they are both useful for traditional social science research. As in all inquiry, however, it is important to triangulate and use multiple methods (Webb, et al., 1965/1999).

## Why Do Evaluation?

There are philosophical, ethical, and pragmatic reasons for peacebuilders to do evaluation.

Expertise is gained not only—or even primarily—through the ingesting of journal articles and manuals or attending trainings, but through *practicing* the ability to practice. Significant research has been done on the nature of expertise, and how experts think differently from novices (Hunt 1994). Experts differ in how they *solve* problems in their domain of expertise, but as importantly, they also differ in how they frame or *set* problems (Schön 1983). Professionals not only develop common ways to solve related problems, but the best professionals continually challenge themselves to question and reexamine their assumptions and their common approaches, sometimes called *double-loop learning* (Argyris and Schön 1974/1992). They critique and grow within a community of practice (Argryris, Putnam, and Smith 1985). The information gathering and the "evaluative thinking" aspect of evaluation support the reexamination of assumptions that allows for improvement. It is the "R&D" of intervention science and designing for change.

In addition to the learning that evaluation provides, any sort of professional practitioner should feel a duty to assess and evaluate the effectiveness of one's work. If your pizza company and car repair shop ask

you to provide feedback so they can better do their job, how much more do those of us who impact people's lives directly have an obligation to request and learn from such feedback? d'Estrée et al. (2001) proposed that while evaluation may present challenges to peacebuilding practitioners, the choice to not evaluate and learn from one's work may be both inefficient and unethical. Can we imagine a physician, a plumber, or a teacher not wanting to determine if their work has produced the desired results?

Finally, there are pragmatic reasons to evaluate one's work. The business of change is inherently messy, probabilistic, and uncertain, and is shaped by the environment. Information gathering allows one to best design an effective intervention, adjust it in process so that it better meets its aims (or recognizes the need for new aims), and at the end of the day identify and demonstrate what difference it made. How better to nurture interest and investment in future work than demonstrating how current work has mattered?

## Why Is Evaluation Resisted? The Importance of a Utilization Focus

We may know people who dismiss or ridicule the value of evaluation. Resistance to evaluation probably comes from two sources: either a misunderstanding of the purposes and forms of evaluation or a past experience with a form of evaluation that was focused on judgment and/or negative feedback. Perhaps we had that coach or piano teacher who was never satisfied, or that boss who loved to check off everything that was inadequate. Organizations may have had past experiences with narrow-minded evaluators conducting narrow evaluations that were perceived as useless at best and harmful at worst.

Evaluation can be transformed into a fruitful experience when the most logical users of the evaluation are engaged in the process of developing the evaluation from the beginning. What information would they like to know? What would be the most useful for their future activities? Patton (2008)

describes *utilization-focused* evaluation as "evaluation done for and with specific, intended primary users for specific, intended uses."

# USING CONCEPTUAL FRAMEWORKS

## What Are Logic Models? Are They Relevant for Our Work?

Though one justifiably might argue that our work in conflict contexts is too complex for a logic model, the task of developing a logic model is an important one and a good conceptual place to begin the planning process. Not only do logic models prevent you from getting away with fuzzy thinking and force you to be specific, but they can also help bring all members of a team or organization on to the same page regarding goals and approaches. I remember working with a peacebuilding organization to sketch out a logic model for their primary initiative and discovering that the executive director and program manager had a different sense of their initiative's goals than the rest of the staff. This is an example of a *process use* of evaluation (Patton 2008; 2012), where evaluative thinking can help cure fuzziness and the operating of disparate and implicit agendas.

A logic model is like a flowchart that captures the logic behind your intervention. The first box leads to the second box, which in turn leads to the third box, and so on. Though labels for these boxes in the flowchart can vary, one standard outline comes from the Kellogg Foundation's *Logic Model Development Guide* (2001), where they label the boxes (or columns) left to right as Resources—Activities—Outputs—Outcomes—Impacts. I personally like to frame the first column as *Inputs* (both Resources and Constraints or Assets and Challenges) (see figure 1.1). The Inputs of the context and organization allow for conducting the *Activities* that in turn have direct *Outputs*. These Outputs lead to desired short-term and long-term *Outcomes*, which in turn beyond the project contribute to the desired *Impacts*.

Several guides for developing logic models (e.g., Kellogg Foundation 2001) will advise you to begin first with developing the Outcomes column, and then fill in the rest of the flowchart. What are you hoping to produce through your intervention? Changes hoped for are often in knowledge, skills, and attitudes (KSAs) but can also be in behaviors, circumstances, status, or functioning (Patton 2012). Outcomes can be thought of as the Goals of your program or initiative. In fact, in the logic model you can see the parallel between the conceptualizing of short- and long-term outcomes, on the one hand, and objectives and goals, on the other.

| Inputs | Activities | Outputs | Outcomes | Impact |
|---|---|---|---|---|
| | | | | |

**Figure 1.1. Logic Model**

## What Is a Theory of Change? Why Is It Important to Articulate This?

Cognitive scientists have mapped out how we develop implicit theories of how the world works that guide our interpretation of it and our behaviors in response. We are never not operating in light of these schemas (or "theories") anytime we engage in any planned, nonreflex action. (Even habituated, automated, and reflex actions are influenced by the schema we apply to interpret events.) Articulating our "theories of change"—how we believe change will happen in our intervention—allows for clarity, on the one hand, and for designing our evaluation as the way to "test" this theory in the field. Theories of change are typically "if-then" statements, such as "If training in conflict resolution techniques is provided, then participants will have alternative behaviors available to use in conflict situations, and will be able to manage conflict in more constructive ways." It is a statement of strategy that articulates why you think your program will work (Kellogg Foundation 2001).

Church and Rogers (2006), drawing on the work of Peter Woodrow and the Reflecting on Peace Practice (RPP) project, summarize many of the assumptions and theories (often implicit) relied upon by peacebuilding programs. These include that peace can be built through changing a critical mass of individuals, developing healthy relationships across groups, withdrawing resources for making war, reducing violence, addressing underlying causes of injustice and threats to identity and security, establishing stable institutions that provide equity and justice, changing the calculus of political elites, mobilizing the grassroots, changing the incentives of individuals and policymakers, and changing public attitudes.

One's "program theory" will often incorporate assumptions from larger tested social science theories of how change happens. For example, improving communication does indeed improve relationship formation (Knapp, Vangelisti, and Caughlin 2014). However, examples exist of where implicit theories of change have actually been disproven by social science research, or where practitioners may at least hold dangerously oversimplified understandings of more nuanced theories about change. Two examples are worth noting here. "Contact theory" is often framed as the assumption that "mere" contact is adequate to improve intergroup relations. However, the actual Contact Hypothesis, first articulated by Allport (1954) and later developed through subsequent decades of extensive research (Pettigrew 1998), shows that intergroup contact leads to improved relations only under certain conditions, and indeed if these conditions are not present, contact can actually make matters worse. These necessary conditions include equal status among groups within the situation; a common goal; intergroup cooperation rather than competition; support from authorities, law, or custom; and the potential to form friendship (Pettigrew 1998). Similarly, many still hold to the folk view that changing attitudes is the best route to changing behavior, when decades of social science research has shown this to be a more complicated relationship (Fazio and Zanna 1978; Kraus 1995; Glasman and Albarracín 2006). Behaviors are influenced not only by more than just attitudes but also by norms, by attitudes toward the behaviors themselves (Ajzen and Fishbein 1980), and by perceived

behavioral control (Ajzen 1991). In fact, structuring situations to change behaviors first will be more likely to have changed attitudes follow.

By making our theories of change explicit, we can verify that our assumptions are valid, and assure that everyone involved in creating the change understands the logic behind the approach taken.

# SHOWING *HOW* WHAT WE DO MATTERS

We may know that our work makes a difference in people and in systems. However, it is important that we are able to demonstrate not only *if* but also *how* things change, and to be thorough enough to capture the good and the not-so-good. How else can we, and our field, learn and improve?

## What Is the Difference between Assessing Outcome and Assessing Impact?

"Outcome" and "impact" are easily confused and conflated. It does not help that these terms are used differently by different funders, agencies, and providers. Church and Rogers (2006) even provide a chart to try to clarify how different terms are used for similar steps in logic models and planning frameworks, noting, for example, that what they consider outcome, the World Bank and UNHCR call "impact." The simplest and most useful distinction may be to consider that "outcomes" are things the program justifiably may be held to account for producing, at least within three to five years (e.g., changes in intergroup attitudes). However, "impacts" are beyond the scope of the project and are usually contributed to by multiple interventions and programs, and thus represent the aspirational goals to which a program hopes to contribute (e.g., regional stability).

The Kellogg Foundation, in their very useful *Logic Model Development Guide* (2001, 2),[1] define impact as "the fundamental intended or unintended change occurring in organizations, communities or systems as a result of program activities within 7 to 10 years, … often after the conclusion of

project funding." Outcomes are defined as "the specific changes in program participants' behavior, knowledge, skills, status and level of functioning," which for short-term outcomes appear within one to three years and for long-term outcomes appear within four to six years.

Proposed outcomes are both short term and long term, and reflect a program's goals and objectives. These latter terms are similarly confused. Church and Rogers (2006, 42) identify these synonyms used for goals (*outcomes*, *project goals*, *overall objectives*, *overall goals*, *development objectives*, and *strategic objectives*) and for objectives (*outputs*, *project purposes*, *purposes*, *specific objectives*, *development objectives*, *operation purposes*, *immediate objectives*, and *immediate results*). However, they consider that a *goal* should be "the broadest change in the conflict that the program hopes to achieve … derive[d] from the conflict or the people in conflict" (30). *Objectives*, in turn, are intermediate steps needed to reach that goal, "the types of changes that are prerequisites" (30). Goals and objectives may be conceived and labeled differently at different points within the same hierarchy, such that, for example, a regional program's *goals* may represent intermediate *objectives* within a larger organization's movement toward an even broader goal.

## In the Complexities of the "Real World," How Can One Ever Say that One's Initiative Has Had Impact? (Attribution versus Contribution)

Given that our world is so complex and multidetermined, it is difficult to establish that the work that we do actually has impact, much less establish any sort of causal linkages in a true scientific sense. We may be able to establish the achievement of desired outcomes as due to program or project activities, but the larger goals we have, defined as impacts, are dependent on changes and interactions among a number of variables, many of which are beyond the influence of even aggregated peacebuilding programs. Classic evaluation approaches, after identifying not only outcomes but also

actual implementation, would use the program theory to attempt to *attribute* the outcomes to the implemented program. However, in many contexts, this level of certainty may be difficult and costly, if even possible, to obtain, and the users of the evaluation results may not require this level of certainty. As noted earlier, Patton (2012) stresses that evaluation design, implementation, and reporting should be done in conversation with the *users* of the evaluation results, those that have a stake in understanding how and if the program works.

For many users, understanding the *contribution* of the program to outcomes is the level of certainty that is both desired and practical. Patton (2012, 355) notes, "Contribution analysis is especially appropriate where there are multiple projects and partners working toward the same outcomes, and where the ultimate impacts occur over long time periods influenced by several cumulative outputs and outcomes over time."[2] First articulated by Mayne (2008; 2011), contribution analysis examines what factors explain an evaluation's findings by weighing evidence and logic against its theory of change, its causal hypothesis. In a manner more like a judicial approach, it uses evidence and argumentation to consider the degree and importance of the contribution, making "plausible associations" and using a standard of preponderance of evidence rather than definitive proof, such that a "reasonable person" would agree. A contribution analysis develops a *contribution story* (Patton 2012) that includes the contributions of the program in light of the linkages in both its theory of change and other influencing factors.

## How Can We Make Sure the Evaluation Will Be Used for Our Agenda Rather than for Someone Else's?

Evaluation is like a ruler; you can use it to measure anything, but the use to which you put that output to can be nefarious or worthy. Patton (2012) argues that evaluation should be done for and with the primary users of the information, which are usually the program's primary stakeholders,

including those designing and implementing the program, those requesting the program, those funding the program, and those participating in the program. The needs of all of these potential information users should be taken into account when crafting the evaluation.

If you are the one crafting and implementing the evaluation for your program, your foremost consideration should be getting you the information to do your program effectively. If you are worried that collecting evaluation data will give someone else the ability, for example, to cut your program, chances are that they will be more likely to cut your program anyway, if you cannot provide information that captures that your program has added value. This is not a reason to resist evaluation. Collecting useful information that will allow you to make program improvements and be responsive to conditions should be in itself a demonstration to sponsors and funders that you care about providing something of value.

## Should I Be Afraid that My Results Will Show I Have Failed?

Elon Musk, the founder of SpaceX, said, "If you are not failing, you are not innovating enough." Gathering information on gaps in services, less than effective activities, and poor attendance *should be* the type of results sought in the formative evaluation stage. As the program is developing, one seeks information for making improvement and for making the program the best it can be. If only great results are being obtained, then one is not fully scanning the environment to see the boundaries of what one is and is not impacting successfully.

During both formative evaluation and summative evaluation, there are many reasons that an intervention may not produce the desired results. *Theory failure* occurs when the logic behind the approach is faulty (see logic model section) or based on untested assumptions such as "shared social media accounts will maintain positive network connections." *Implementation failure* occurs when the intervention has not been implemented in the way the model had been originally envisioned, such as

in an incomplete way, or only half the time it was supposed to be, with inadequate strength, and so on. For example, conflict resolution training in schools for only thirty minutes once a week may not produce noticeable constructive results, not because it is not useful or an effective intervention but because it is not *enough* of the intervention to see change.

## USING EVALUATION TO IMPROVE CLARITY AND REFLECTIVENESS

The *process* of doing evaluation has its own benefits, apart from producing findings that can be used for decision-making. These can include clarifying goals, getting all staff and other stakeholders on the same page, reinforcing interventions, and empowering people to engage in their own learning. The skill development and increases in learning that come from the evaluation process adds another layer of value beyond the findings themselves.

One way to clarify goals is to engage in an *evaluability assessment* (EA) (Smith 2005). An EA basically determines if your program is ready (i.e., clear enough) to be evaluated in the traditional formative and summative ways. It includes clarifying goals and central issues, identifying the expectations of relevant stakeholders, and making the model that is going to be evaluated explicit and clear. The process of doing evaluation requires that these be clear first, but this clarification process is itself extremely valuable. An organization can determine how much more thinking needs to be invested before they are really ready to implement effectively, staff and leadership can make sure they all share a common understanding of goals rather than muddled assumptions, and disappointed expectations can be avoided from the outset. The thinking needed for such clarification is a welcome and refreshing exertion on the front end that will both make the intervention more effective and make the evaluation flow more naturally.

We all agree that increasing reflection would be a good thing. The process of evaluation can improve reflection both for participants and for the organization doing the intervention. For participants, reflection on their

learning could likely reinforce whatever changes in knowledge, attitudes, or behaviors were originally desired. This same reflection record can at the same time be used for evaluation purposes. For example, Patton (2012) describes how data collection might be integrated into a workshop by asking participants to begin with an exercise where they reflect on their existing knowledge, skills, and attitudes. According to adult learning principles, this can help prepare them for the upcoming learning in the workshop. Then, at the end of the workshop they can be asked to retake the self-assessment they began with, in order to assess and reflect on what they have learned. Again, according to learning principles, this can reinforce the learning that has taken place. These two reflection exercises not only reinforce learning but also conveniently provide a pretest and posttest for evaluation without ever having to name it as such.

Shared understanding can be fostered by broadening who is involved in deciding how change will be tracked. Participants can help to develop meaningful measures (see Rothman and Sachare, this volume; Firchow and Tilton, this volume), and can also be taught evaluation methods. The process of defining what matters can give people voice and allow them to learn others' perspectives.

All aspects of evaluation can be used to increase the reflectiveness of the organization doing the intervening. However, building evaluation practices and logic into standard operating procedures can make all choices and decision-making more reflective. This can include staff learning evaluation methods, and learning how to see things differently. Evaluation can become part of an organization's culture, such as supporting and normalizing data-informed decision-making and institutionalizing forms of sharing and learning best practices. Evaluation can be used not only to improve a program but also to *develop* a program. This newer form of evaluation, called *developmental evaluation* (Patton 2011, 2012) uses evaluation approaches to inform organizational and program development in real time, as it adjusts to a complex and evolving context. Evaluation is used not just to improve intervention but to continually recalibrate and rethink the model itself.

# NUANCES AND NETTLES

These are questions that come from those who know enough to want to know more. I have seen people try to finesse these questions, which then adds to questioners' confusion and alienation from evaluation rather than enlightening them. Here are short answers to topics that really require more space, but with references for those who want to understand more (and definitely for those who are asked to teach others on these topics!).

## If We Use Evaluation to Reinforce an Intervention, Are We Compromising the Evaluation's Objectivity?

Using evaluation processes to reinforce the intervention does pose threats to validity from a methodological perspective. However, if the purpose of information gathering is ultimately to increase the likelihood of the desired change that the intervention is seeking, then this can happen both through gathering information to improve the intervention *and* through using the information gathering to strengthen the intervention (see section on increasing reflection, above). If the users of the evaluation have as their primary goal the achievement of change, then "as long as [e.g.,] the instrument items are valid indicators of desired outcomes and the desired outcomes are being obtained" (Patton 2012, 150), it may not matter how much of the change is from a pretest or posttest sensitization or from the activity itself. Moreover, the integration of the data collection into the intervention means it will take place as a central component of the intervention and not as an afterthought or additional burden.

## Do I Need a Control Group?

Methods debates go on in all fields, including in evaluation. Ultimately your methods should serve the purposes of why you are doing the analysis. If you are trying to prove a direct causal relationship between your intervention and a certain outcome, as one might in physics or

pharmacology, then you will need a counterfactual—a control group that does not receive the intervention and yet is matched to your treatment group in all the important ways including proceeding along the same trajectory. However, developing this sort of test means that (1) you are able to standardize your intervention to give it a strong test; (2) you are able to hold all other factors constant to establish the causal influence of your intervention; and (3) the next time this intervention is applied in order to seek the same result, the same surrounding conditions will be in place. This works well for simple interventions, in controlled environments. It is seldom appropriate or desirable in the complicated and dynamic contexts of peacebuilding.

If you are doing your analysis in order to learn if a credible case can be made that your intervention has contributed to some change, then building in a control group may be a poor choice of resource use and may be politically difficult and/or ethically problematic if you are withholding "treatment." The question mainly revolves around what level of evidence is needed for decision-making, which is the primary purpose of evaluation, and what trade-offs must be made in order to achieve the level of evidence that is sufficient for that decision-making and those decision makers. This choice should be made in discussion with the primary users of the information. As noted earlier, there are fields where "a preponderance of evidence" is the standard, rather than "certainty at the $p \leq .05$ level."

## What Are "Randomized Control Trials"?

Randomized Control Trials, or RCTs, are considered by a subset of researchers to be the "Gold Standard" of the field of evaluation, meaning evaluation as used in many diverse forms of intervention, from pharmacology to education. They involve randomly assigning participants to conditions, such that any participant has an equal chance of being in a control or a treatment group, and there is no systematic bias in who receives what. This helps one to be able to say that the groups before the intervention are equivalent, and thus any later differences must be

explainable as being due to the intervention itself. This favoritism for RCTs holds experimentation to be the best form of gathering evidence and drawing conclusions.

Most social scientists, even those who come from disciplines where experiments are common (like my home field of psychology), would never say that experimentation is the only or even the best way to learn things about the world. It is true that experiments are considered the best way to achieve internal validity, that is, establish causal relationships between the variables studied. However, experiments are weaker on external validity, that is, going beyond the particulars studied in this setting to less controlled, more naturally occurring settings.

RCTs are able to answer if one particular intervention in a particular setting caused a particular outcome. However, what is often of more interest to those making decisions is what are the relative impacts of different interventions, or different interventions for different people. Patton (2012) suggests that the type of most *useful* information to gather should be the "gold standard" for deciding upon the method (Patton 2012).

Patton suggests that "experimental designs can be appropriate where the program is conceived of as a simple intervention that yields discrete, identifiable outcomes through a process of bounded and linear cause-effect relationships" (2012, 298). However, two things should be borne in mind. The first is the stage of the program's development and the purpose of the evaluation. An experimental design would most likely be used only for the final, summative evaluation, once a program has been developed, tested, and sufficiently evolved and improved to be considered stabilized (see earlier formative versus summative distinction): It "would be inappropriate to impose an experimental design on a new or innovative intervention before it has undergone a period of formative evaluation" (298). The second thing to bear in mind is the complexity of the situation and the stability or dynamism of the intervention needed. In highly dynamic, unstable, emergent settings—which characterize most peacebuilding settings—intervention may need to be constantly responsive and evolving (see

d'Estrée, this volume; Woodrow and Jean, this volume), thus making the control needed for experimentation inappropriate and impossible to achieve.

## What Are Some of These (New, Innovative, Unusual, Boutique, Alternative) Forms of Evaluation Such as Empowerment Evaluation and Action Evaluation?

Empowerment Evaluation (Fetterman and Wandersman 2005) suggests that evaluation is not only for answering decision-making questions but also for increasing the ability of people and communities to gather the information they want and need in order to change their own lives. Like other forms of empowerment, it seeks to increase both personal and community self-determination. It moves the ownership of the evaluation to the parties themselves and places a primary emphasis on building capacities and continual and ongoing learning.

Similarly, Action Evaluation (Rothman 1997a; see also Rothman and Sachare, this volume) focuses on determining the goals of the evaluation from the parties themselves through a sort of consensus-building process and builds in ongoing learning and adjustment of goals and objectives as activities progress and the context evolves.

## What Is Participatory Evaluation? Is This Participatory Action Research?

Participatory Evaluation changes the role of the participants from one of participating in the evaluation to cocreating the evaluation. The evaluator plays more of a facilitative role, and any status differences are minimized (see Neufeldt, this volume). Participants are encouraged to take on the evaluation themselves with the support of the evaluator, including developing the measures (see Firchow and Tilton, this volume), engaging in data collection, and reflecting on the results. Participants learn evaluation skills and logic and determine the focus of the evaluation that matters to

them. Participants learn that their views are valued, and they learn to value each other's views, working as a collaborative group. Accountability to any outside audience is secondary to the accountability participants take on to themselves and their community.

Participatory Evaluation can be considered a form of Participatory Action Research (PAR). PAR has been around since the 1940s. It is a collaborative approach to research where the goal is to learn about the world by trying to change it. Researchers and participants collaborate to plan and implement change, collect data, and reflect on the results. One of the primary areas in which it has been applied is education, but it has also been applied to various forms of development. It includes both concern for research and the advancement of knowledge, as well as for social transformation and empowerment of marginalized people.[3]

## CONCLUSION: TRAINING FOR OPENNESS

Evaluation is at its essence about learning. As peacebuilding practitioners we rely on experience, research and theory-informed practices, and wisdom from our professional communities and our local partners. Evaluation also prepares us to be continually open to new input and new learning and to approach each new and unique situation with humility and "beginner's mind." Understanding the many concepts and tools provided by evaluation methods and processes further trains our intellectual muscles to be prepared to be reflective and responsive in facing the challenges along the complex paths of peacebuilding.

## NOTES

1. Available at https://www.wkkf.org/resource-directory/resource/2006/02/wk-kellogg-foundation-logic-model-development-guide. Accessed Sept. 10, 2018.
2. See also Jones, this volume.
3. For additional information, see Reason and Bradbury (2008).