© OECD 2012

Chapter 4

Conducting an evaluation in situations of conflict and fragility

Chapter 4 considers the business of conducting an evaluation. It begins with the inception phase, and then looks at how to identify theories of change and the implementation logic underpinning the activity being evaluated. The next step is the issue of gaps in baseline data and how and where to source data in order to plug the gaps. The chapter looks at the criteria evaluators should use, focusing in particular on the DAC criteria of relevance, effectiveness, impact, sustainability and efficiency. It then describes how to bring an evaluation to a close. The chapter looks at drawing conclusions and issuing recommendations and at the reports evaluators produce. The next step the chapter discusses is communicating the evaluation's (positive or negative) results to stakeholders and disseminating the lessons learned. Finally, to close the loop, the chapter emphasises the importance of feeding findings back into programme design and management.

A number of core steps should be part of any evaluation. Evaluators and managers (and others involved in the actual process of evaluating) do not have to perform them in the order in which they are set out below. But they should bear them in mind.

Key steps in the evaluation process

- ✓ Allow for an inception phase
- ✓ Identify the implementation logic and theory of change
- ✓ Deal with missing baselines and other gaps
- ✓ Gather data
- ✓ Examine the effort using various criteria
- ✓ Draw conclusions and make recommendations
- ✓ Conduct reporting
- ✓ Ensure quality
- ✓ Feedback on the evaluation
- ✓ Management response
- ✓ Disseminate findings
- ✓ Feedback into programming and learning

Allow an inception phase

Given the complexity of conducting evaluations of donor engagement in fragile and conflict-affected situations and the underlying weaknesses in programming, an inception phase can help identify issues that need to be addressed before proceeding with the evaluation. The scope of an inception phase can range from simple to complex. Usually it will involve desk study, document review, and the production of an inception report, but some preliminary field work may also be required. An inception phase may be useful for assessing or conducting conflict analyses. It could also be used to map the evaluation subject, conduct a donor policy and portfolio or country policy analysis, or carry out thematic studies as needed.

Evaluators present the results of an inception phase in an inception report that can be the basis for discussing data availability, evaluability, and the feasibility of planned data collection strategies. Inception reports are also useful tools for adapting methods and for fine-tuning the approach chosen to address the key issues specified in the terms of reference. (See "Reporting" in this chapter).

Identify and assess the theory of change and implementation logic

A theory of change is a set of beliefs about how change happens. It explains why and how people think certain actions will produce the changes they desire in a given context, at a particular moment in time (Weiss, 1995; Church and Rogers, 2006). It is a term used to

describe the links between the context, the intervention inputs, the implementation strategy, and the intended outputs and outcomes. Implementation logic, also called programme logic, is a term that describes why an activity is doing what it is doing, usually at the project level. It is closely linked to the theory of change. Evaluation can reveal whether success, or lack of it, is attributable to programme design and theory or to implementation. Contributing to testing theories of change and implementation logic is one of the prime contributions evaluation can make to research and broader learning.

In situations of conflict and fragility it can be especially important for evaluators to identify theories of change, since they are often implicit, unexamined, and untested. In Kosovo, for example, the international community operated for several years on the assumption (theory of change) that peace could be achieved by improving relations between the two main conflicting parties. On the basis of this theory, it funded numerous programmes to promote dialogue, exchanges, youth interaction, women's groups, and so forth. All were aimed at building cross-community relationships. However, a study found that without intra-community work to bridge internal divides and create more responsible leadership, cross-community interventions had little effect. The activities were operating on an incomplete theory of change. The study also found that cross-community relationships did not help prevent violence or strengthen collective resistance to violence in 2004. Bonding social capital – intra-ethnic networks of trust and reciprocity – and mono-community leadership, motivated by the strategic desire to secure independence for Kosovo, led to more effective action (Chigas *et al.*, 2006).

Causation: Between theory and outcomes

A theory of change can take different forms at different levels. Theories of change at the country and conflict levels may be quite general – for example, strengthening the capacity of the government will help improve governance and therefore reduce conflict and violence. In this broad vision, development actors may operate in different ways (e.g. through budget support or capacity support to individual ministries) or work in different sectors, such as social services or finance and planning.

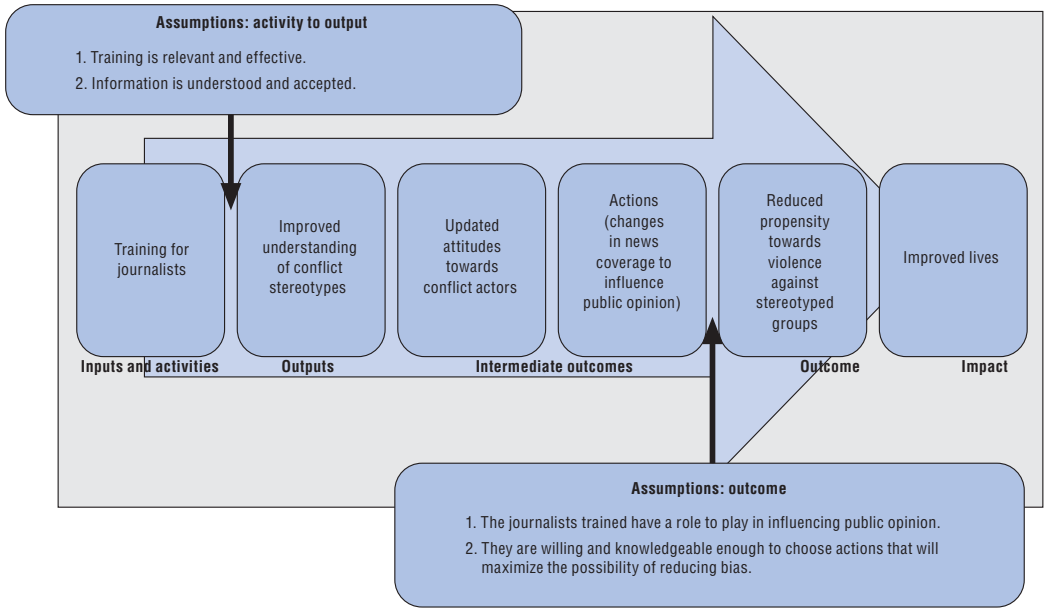
Evaluations of specific projects or programme would be likely to formulate more directly causal theories of change which link specific inputs and activities to desired micro-level outcomes and to broader peace dynamics. For example, in the Southern Sudan evaluation (Box 3.6), the team identified broad theories of change across the full portfolio of interventions they were examining. They did not drill down to the specific implementation logic level or look at individual projects. Instead, they grouped the multiple activities of different donors under broad theories of change, and then tested each one.

Programme documents do not always state in explicit cause-effect terms how a programme aims to produce its intended outputs, outcomes, and impacts. Accurate, clearly worded theories of change are necessary for effective programming and should be evaluated. More importantly, the assumptions that underpin theories should be the subject of evaluation. By identifying how an intervention was expected to contribute to sustainable peace or address conflict and fragility factors, evaluators use the theory of change to assess relevance, effectiveness, and impact. An example of this thinking is the imaginary anti-bias peace journalism programme described in Annex C and illustrated in Figure 4.1.

The programme could be understood to work on a theory of change whereby it will train journalists and increase their knowledge of conflict dynamics, so reducing bias in

reporting and contributing to a reduction in tensions and violence between warring groups. An evaluation of this programme would test the underlying theory that improved knowledge reduces reporting bias and the secondary theory that improved reporting lessens the tendency to resort to violence in the broader community. Key questions for analysis would be how workshops, awareness raising, and skills development actually change conflict reporting and what impact they have on critical conflict dynamics. The programme could track the language used in reporting before and after training. It might also survey public attitudes and, at the same time, the programme activities to see if they were achieving the expected results or whether unexpected obstacles had arisen. It might turn out, for instance, that individual journalists have very little influence over the use of inflammatory language and that editors determine the use of "colourful" language to boost sales, so reinforcing stereotypes. Such a finding would suggest that the "theory" of training journalists to influence public opinion was flawed.

Figure 4.1. **Assumptions underlying the theory of change in a fictitious peace journalism programme**



Source: Content from OECD (2008b); diagram adapted from Gaarder, 3iE (2011).

Where do theories of change originate?

One task related to identifying and evaluating theories of change is to identify where they originate. Are they empirical – based on solid evidence from past programmes? Are they drawn from programme designers' personal and professional experience or from the experience of the stakeholders and beneficiaries consulted during programme design? Or are they research-based?

When working theories are not explicit they will need to be uncovered through document review or interviews with implementing staff and other stakeholders. The evaluation process may also reveal that different stakeholders involved in a particular intervention operate on different assumptions (theories) about how their efforts will promote change for peace. By drawing out underlying theories and opening up a discussion of the validity of different theories, the evaluation process can be useful in helping those

involved in a programme reach consensus on what they are doing and why. This clarity can help improve programme design and implementation.

Evaluations of donor engagement in fragile and conflict-affected situations very often cover multiple projects and – in some cases – multiple donors. Such donor engagement may involve multiple theories or multiple variations of a broad general notion, e.g. more aid for social services supports stability.

Box 4.1. **Evaluating success and failure in peacebuilding**

When evaluating it is important to distinguish between and analyse various types of success and failure as an input to learning and future programme design. In this field it helps to distinguish a failure of the theory (wrong assumption about how change will happen) from a failure in implementation. Theory failure indicates the failure of a conflict prevention or peacebuilding activity due to a flawed causal relationship – in other words, underlying assumptions about how to bring about change in this context are false. A faulty theory of change could be based on an inaccurate conflict analysis, or it could reflect misdirected priorities or mismatched objectives.

Implementation failure denotes a problem with the execution of the activity itself (inputs/outputs, staff capability, timing, location, security environment or budget) or with management systems. Such problems could include sudden changes in the conflict that disrupt or reverse progress, despite an otherwise well designed activity.

Source: OECD (1999).

Gather data

As mentioned in the introduction, evaluators and evaluation managers sometimes encounter weaknesses in policies, strategies, and interventions. There may be unclear or unstated objectives, a poorly articulated theory of change or programme logic (as described above), missing indicators, no monitoring data, or no baseline information. The security situation can have serious implications for data and may particularly affect the ability of evaluators to travel to certain regions or countries and to access affected communities, programme beneficiaries or other key informants.

Evaluation teams, working in co-operation with the commissioning organisations, must address the issue of weak or missing data. They should consider ways to (re)construct or compensate for missing baselines and other data during the evaluation process, bearing in mind that nothing can ever fully replace solid planning. Tips on how to compensate for or work around such gaps without endangering the quality of the evaluation can be found in a number of publications (World Bank, 2005b; Bamberger *et al.*, 2006; OECD, 1999). Box 4.2 considers an example of how a donor-supported project in Afghanistan is working to increase the availability and quality of data.

A key phase in the evaluation process is collating data to trace the "story" of a programme or policy and its effects, which include how they affect stakeholders' viewpoints. Evaluators gather information from programme documents and reports, any available monitoring data, and interviews with programme staff, partner organisations, local officials, target groups, participants, third parties (groups such as neighbouring communities not targeted by a programme), and international and national actors who

Box 4.2. **Building statistics for monitoring and evaluation in Helmand Province**

The Helmand Monitoring & Evaluation Programme (HMEP) is a programme funded jointly by the UK's Department for International Development (DFID) and the Provincial Reconstruction Team (PRT). It is designed to assist improved delivery and effectiveness of stabilisation and development programming in Helmand, Afghanistan. HMEP collates data from third-party sources as well as drawing on a dedicated research capacity to establish baselines and monitor indicators of progress against the Helmand Plan. Data is presented graphically and geospatially and stored on an interactive database, accessible from an online website. HMEP produces quarterly monitoring and analytical reports and up to four *ad hoc* reactive reports per year aligned with PRT and DFID reporting requirements. In addition, HMEP is set up to support the development of programmatic capacity in the PRT and the integration of monitoring and evaluation into planning and programme implementation.

Source: Coffey International Development, "Helmand Monitoring and Evaluation Programme", *http://uk.coffey.com/our-projects/helmand-monitoring-and-evaluation-programme*.

observe the intervention from a distance (civil society organisations, research institutions, donors, media, academia, think tanks).

Data collection strategies depend on an evaluation's design, its purpose, and the information sources available. Strategies that can be used include random and purposeful sampling. Quantitative and qualitative data can be collected through censuses, observation, household surveys, interviews, questionnaires, anthropological or ethnographic research, participatory workshops and discussion groups. National statistics systems and major NGOs will often have available demographic data, though these may not be suited to the required sample size or detail. Moreover, some private polling companies have begun to specialise in household surveys and data collection to professional standards in conflict-affected areas. Box 4.3 gives examples of data sources and the evaluation literature provides further detail on data collection techniques.

Prolonged violence and situations of high tension pose significant data availability problems that often restrict evaluators' work. However, there might be more data available than is first evident in the planning phase and in programme documents. High staff turnover often shortens institutional memory. Moreover, different external and national actors are frequently poorly co-ordinated and do not always share the information they collect with other actors to whom existing studies, assessments, data, evaluations and surveys may not be known or available. Valuable baseline data might be gathering dust in documentary archives somewhere with few aware of it. It is sometimes worthwhile investing time up front in an evaluation phase to map data already collected, either as part of the evaluation inception phase or as a pre-study.

Evaluators and other researchers should share and use data collected by others, as too many are gathered and allowed to lie fallow. It is also advisable that those commissioning an evaluation and the evaluation team agree to contingency plans for sudden shifts in the situation on the ground that may impact collection of data that are necessary for the evaluation. Evaluation managers may include data management provisions in the terms of reference.

Box 4.3. **Examples of data sources from Somalia and Afghanistan**

Somalia: Using multiple data sources and accounting for data shortcomings

A 2011 evaluation conducted by DARA (an independent organisation specialised in humanitarian evaluation) looked at the effectiveness of the humanitarian response in Somalia. It drew on the following data sources:

- Literature on past assistance to Somalia and the contextual variables, and 300 web pages and relevant publications, and assorted documents.
- Semi-structured group and individual interviews with 489 stakeholders. Of these, 189 (112 in Kenya and 77 in Somalia) were carried out with a range of individuals involved in the response, who included representatives from UN agencies, the Red Cross and Red Crescent, international and national NGOs, local government actors, and donors. Women accounted for 24% of interviewees. There were group interviews with over 300 people from the affected population.
- Field observations in Kenya and Somalia, in camps for the internally displaced and Somali refugees.
- Online survey of former staff and key informants involved in the humanitarian response (response was low with only ten respondents answering).

In total, the evaluation team gathered 3 117 items of information but encountered the following challenges: high staff turnover, limited time for field work, insecurity, poor consistency of data, disaggregation gaps (limited availability of data disaggregated by age, gender, vulnerabilities and outcomes in the period under review). Safety concerns prevented international evaluators from visiting areas proposed in the terms of reference, although national evaluators were allowed to visit some.

North East Afghanistan: Creating new data

The more fragile a state is, the less data tend to be available. Evaluation teams often have to collect their own data. In the context of an evaluation of the impact of aid in North East Afghanistan, the evaluation team gathered data from various sources: The core of the data was collected with two surveys among 2 000 head of household respondents in 80 villages. Evaluators also collected data from other sources to create a profile for each community, containing information on the history, demography, ethnic composition, political and social organisation and resource endowment. In other settings, much of these data would be readily available from existing sources, such as a national census. In the Afghan context, however, evaluators had to collect these data themselves – which was possible only in the context of a multi-year research project. In order to collect information on major events and changes affecting the communities, a monitoring and reporting system was set-up covering 40 villages. Local correspondents completed semi-structured reports four times a year. The evaluation team collected additional qualitative data during field research visits. Finally, evaluators obtained data on aid flows from various international development organisations in the region. The data were then compiled and used for quantitative and qualitative analysis of the impact of aid on stability.

Sources: Polastro *et al.* (2011) and Böhnke *et al.* (2010).

During conflict and in its immediate aftermath, when mistrust is rife and most intervention stakeholders also have a stake in the conflict, the reliability (not to mention availability) of data and information provided is often particularly problematic. Depending on their own position within the conflict, different actors may have diverse, or even

contradictory, interpretations of an intervention's positive and negative impacts or relevance. To ensure reliability, evaluators should use multiple sources or types of information and a mix of sound quantitative **and** qualitative data. They should triangulate the data they use, ensure that sources are transparent, and verify the data's validity before analysing them – by fact checking with key stakeholders and interviewees, for example. By combining multiple data sources, evaluators can offset the bias that comes from relying on a single type of information and single observers. In the evaluation report's description of methods, any issues around data (including data gaps or problems with inconsistency) should be described along with the impact these data problems had on the reliability and validity of the evaluation's conclusions.

To avoid increasing tension between groups (in accordance with the "do no harm" principle), decisions about how to involve various groups in data collection should be based on a clear understanding of stakeholder roles and interests, drawing on the understanding of stakeholders developed in the conflict analysis. Where access or security concerns impinge on data gathering, other methods – including consulting with knowledgeable proxies able to provide representative views or information – could be explored. Consideration to coverage may be required – to include perspectives from both within and outside the capital city, for example – depending on what issues the evaluation intends to address.

Evaluations should draw on accurate, relevant data about women, men and gender relations. This helps ensure evaluations are gender and conflict sensitive, makes gender disparities more visible, and assists in answering the evaluation questions. Gender-disaggregated statistical data and gender analysis (information on the position of women and men in a particular conflict context) should inform the evaluation. Where disaggregated data are not available, combinations of quantitative and qualitative approaches can be used to gain the necessary understanding of gender relations and differentiate results for men, women, boys and girls. Where it is deemed that inclusion of such data is not necessary or not feasible, this should be clearly justified in the evaluation report. For example, in the case of an evaluation of the impact of aid on local perceptions in rural North East Afghanistan (Böhnke *et al.*, 2010) the evaluation team used a survey. Though the team wanted to include women and youth, in the end the survey respondents were all male heads of households. Their decision was based on an analysis of the local context and consultation with regional and gender experts, and on an earlier experiment using female researchers to survey Afghan women, which found that security threats and the social custom of having men accompany each female researcher made it nearly impossible for the teams to work effectively. The evaluators decided that the choice not to include women in the survey was justified in this case because the opinions of older men (as the political representative of their household) were most relevant to answering the key evaluation questions. They verified the decision with experienced local teams and tested it before doing the initial baseline. In the evaluation report the team explained this decision, stating that, "while we think that the tasks at hand warrant these choices and the resulting limitations, we are nevertheless conscious of the fact that our research design is not equipped to capture trends for all of Afghanistan, nor does it capture perceptions of women [...]" (Böhnke *et al.*, 2010). Transparency over methods and data limitations is a requirement of high-quality evaluations.

Many interventions work to build peace and prevent conflict by creating change in people's attitudes, thought processes, and relationships. In such cases, it may be necessary

to collect attitudinal data, conduct interviews, workshops, or focus group discussions with stakeholders, or carry out surveys to collect quantitative data. Measuring intangible changes in areas such as perceptions through interviews requires the same triangulation vetting as other types of data.

Relevant interviewees (primary sources) may be missing, because they are either in prison, dead, or unavailable in other ways. The evaluation report must account for such data gaps and make clear how they have been compensated for. Investing time early in the preparation phase for researching the data situation should help prevent surprises like missing data at a later stage, which can derail or delay the evaluation process.

Criteria for evaluating interventions

When evaluating development co-operation programmes and projects it is useful to consider the criteria outlined in the *DAC Principles for Evaluation of Development Assistance* and additional criteria that may be valuable. The analysis of these criteria forms the main content of an evaluation report. This section suggests how each OECD DAC evaluation criterion might be adapted to the field of conflict and fragility. The specific questions outlined below are guiding examples, not a comprehensive or obligatory list. The evaluation criteria are interlinked, each criterion shedding light on the intervention being evaluated from a slightly different perspective in order to develop as comprehensive a picture as possible of the intervention. When read together, the criteria should assist the evaluation team in developing a clear understanding of the activity or policy being evaluated and its contribution to statebuilding and peacebuilding. This adaptation of the criteria is based on contributions from Paffenholz and Reychler (2007), the draft guidance (OECD, 2008b), OECD and CDA (2007), and Anderson and Olson (2003).

Relevance

The relevance criterion is used to assess the extent to which the objectives and activities of the intervention(s) respond to the needs of beneficiaries and the peacebuilding process – i.e. whether they address the key driving factors of conflict revealed through a conflict analysis. Relevance links the outcomes of the conflict analysis with the intervention's objectives, although the relevance of the intervention might change over time as circumstances change. Understanding relevance may also involve an assessment of the extent to which an intervention ties in with overall strategies and policy frameworks of the country or external partners. Different conflict groups or actors may have different perspectives on the relevance of an intervention and its results. Women and men may also perceive the relevance of the intervention differently.

Assessing an intervention in relation to the conflict is key to evaluating its relevance. If staff, managers, or others involved in design and implementation have already carried out some kind of conflict analysis, its accuracy and use should be assessed. Assessing whether or not the conflict analysis proves (or has proved) accurate will be an important aspect of the evaluation. Assessing the analysis will contribute to learning and to refining theories about why violence occurs and what the most important determinants of conflict dynamics and long term statebuilding are. If no process of systematic analysis has taken place previously, the evaluation team may either talk to staff and stakeholders to understand what underlying (unarticulated) conflict understanding is guiding their work, or facilitate a more formal exercise for conducting a conflict analysis. The analysis is

needed to identify the expected connections between the programme outcomes and peacebuilding goals and the relevance to key driving factors of conflict and fragility.

Questions on relevance might include the following:

- Is the intervention based on a valid analysis of the situation of conflict and fragility? Has the intervention been flexibly adapted to updated analyses over time?
- In the light of the conflict analysis, is the intervention working on the right issues in this context at this time? Does the intervention appear to address relevant key causes and drivers of conflict and fragility? Or does it address the behaviour of key driving constituencies of the conflict?
- What is the relevance of the intervention as perceived by the local population, beneficiaries and external observers? Are there any gender differences with regard to the perception of relevance?
- Are the stated goals and objectives relevant to issues that are central to the situation of conflict and fragility? Do activities and strategies fit objectives, i.e. is there internal coherence between what the programme is doing and what it is trying to achieve? Has the intervention responded flexibly to changing circumstances over time? Has the conflict analysis been revisited or updated to guide action in changing circumstance?

Effectiveness

Effectiveness is used to evaluate whether an intervention has met its intended objectives with respect to its immediate peacebuilding environment, or is likely to do so. The key to evaluating effectiveness – and thus the linkage between outputs, outcomes and impacts – is finding out to what degree the envisaged results have been achieved and noting changes that the intervention has initiated or to which it has contributed. Furthermore, as most of the activities undertaken in situations of conflict and fragility are not explicitly geared towards sustainable peace, it is important to draw a distinction between two kinds of results. One is "programme effectiveness", i.e. to what extent the programme achieved its stated objective. The other is – if the programme met its objectives or goal – the immediate or secondary outcomes as they relate to peacebuilding and conflict dynamics identified in the analysis.

A theory of change analysis can be used to assess the criteria of effectiveness. This involves assessing whether an intervention is based on a sound theory and logic and whether these are proving (or have proven) to be true. The understanding of effectiveness is also linked to the conflict analysis. A programme or policy may do good or do well and still not change the underlying dynamics or key driving factors of conflict and fragility identified by the conflict analysis. Also, external factors, unrelated to and beyond the control of the activity in question, may be more significant factors of peace and conflict, in which case effectiveness will be understood relative to these broader dynamics and trends.

Box 4.4 describes how Germany evaluated the effectiveness of its civil peace programme using a framework that combined theories of change analysis with empirical research.

Box 4.4. **Evaluating the effectiveness of the German Civil Peace Service**

Germany commissioned an independent evaluation of its Civil Peace Service programme, which promotes non-violent ways of dealing with conflict through a variety of education, training and service activities. The evaluators developed a comprehensive evaluation framework that enabled systematic comparison of data across eight country cases. To compensate for insufficient data, a combination of evaluation approaches and methodologies was developed. In addition to reconstructing baselines and describing the Civil Peace Service's underlying theories of change, the evaluation team assessed the plausibility of outcomes, i.e. the likelihood that the intervention had achieved, or would achieve, the hoped-for outcomes. This was done by checking Civil Peace Service interventions against a list of conditions for effectiveness developed during empirical research. The evaluation team also identified good practice examples from among the interventions evaluated and then identified which factors contributed to their effectiveness.

Source: Paffenholz (2011).

Examples of questions that can help determine effectiveness include:

- Has the intervention achieved its stated (or implicit) purpose, or can it reasonably be expected to do so on the basis of the outputs and outcomes?
- Is the theory of change based on valid/tested assumptions? Are there alternative theories of change?
- To what extent did donors identify and adequately manage context-specific risks?
- Is the intervention achieving progress within a reasonable time frame, or will it do so? Is it possible to accelerate the process? Should the effort be slowed down for any reason?
- What major factors contribute to the achievement or non-achievement of objectives?
- Has the intervention achieved different results for women and men and boys and girls?

Depending on the activity being evaluated and the scope of the evaluation, some specific examples of conflict-specific questions on effectiveness include: Does the intervention prompt people to increasingly resist violence and provocation? Do the stakeholders affected have a significant impact on the situation of conflict and fragility? Are the right/key people or many people being addressed? Were gender equality and relevant horizontal inequalities (ethnic, religious, geographical, etc.) that drive conflict taken into consideration and what are the results? Does the intervention result in an increase in people's safety and in their sense of security? Does it improve non-violent forms of conflict resolution or power management? Does it result in a real improvement in relations among groups in conflict, demonstrated in behaviour changes?

Impact

The criterion of impact refers to the wider effects produced by an intervention. Such effects may be positive or negative, and may be produced directly or indirectly, intentionally or unintentionally. In fragile and conflict-affected contexts the criterion of impact is used to identify and evaluate the effects of the intervention on the key driving factors and actors of the conflict, as well as on broader development and statebuilding processes, as relevant. Assessment should cover both the desired changes the intervention aimed to achieve and any unintended (or unexpected) positive or negative results.

Changes in behaviour and attitudes, of the kind many peacebuilding interventions seek, are often difficult to measure and can take a long time to achieve. With this in mind, it may be too soon to reasonably expect significant impacts on conflict drivers (for example, for activities aimed at reforming institutions). In this case the evaluation might focus on outcomes and short term impacts and test the theory and programme logic to predict whether the current strategies are reasonably likely to contribute to peace over the long run.

Methods for evaluating impact are covered elsewhere in evaluation literature, for example in the guidance commissioned by the Network of Networks for Impact Evaluation (Leeuw and Vaessen, 2009). The rigorous quantitative methods associated with impact evaluation and randomised control trials are considered not feasible in many situations of conflict and fragility, (although useful experiments are underway at the programme level). Still, it is particularly difficult to apply such methods to high-level questions of peace and conflict across various interventions at country level or to assessments of overall donor engagement in a conflict setting. Where causality cannot be reliably determined using rigorous methods, evaluators may present plausible explanations for their conclusions regarding impact, though limitations should be made explicit. Box 4.5 describes quantitative impact evaluation methods in more detail.

When violent conflict is extreme, analysis may have to shift towards monitoring or "real-time" evaluation functions that focus on output indicators. They are easier to track and less prone to attribution failures than impact indicators.

Examples of questions that can help determine impact may include:

- What are the primary and secondary, direct and indirect, positive and negative, intended and unintended, immediate and long-term, short-term and lasting effects of the activity or policy in question? Does it exert a significant effect on key factors for conflict or peace?
- Drawing on the conflict analysis, what key drivers of conflict and fragility were affected and how? What changes can be ascertained in attitudes, behaviours, relationships or practices (of how many people and/or classified according to selected criteria such as gender)? Are there any secondary negative effects?
- How has the situation changed over time and what, if any, has been the contribution of the intervention to those changes?
- What impacts have the interventions had on specific indicators of well-being, such as health status or poverty levels, addressed by the intervention? What are the impacts on long-term development trajectories?
- Has the intervention impacted policy? How do these policies relate to the conflict?

Sustainability

Sustainability is defined as the continuation of benefits on end of assistance. In an environment of conflict and fragility, sustainability includes the probability of continued long-term benefits and resilience to risk over time, as well as lasting benefits in the economy, institutions, human resource management, etc. As in other fields, sustainability also includes "ownership" of the peace and development processes. Experience and peace research demonstrate that peacebuilding processes are long term and call for long-term engagement that can weather setbacks (OECD and CDA, 2007). In conflict-affected regions,

Box 4.5. **Quantitative methods to evaluate impact in settings of conflict and fragility**

An impact evaluation measures the net impacts of an intervention by comparing its results with a counterfactual – a measure of what would happen in the absence of the intervention. Although experience with the impact evaluation approach in peacebuilding interventions is limited, it is being used increasingly in the development context and, indeed, in conflict and fragile settings.

There are two main methods of conducting a counterfactual comparison. The first is the experimental approach or randomised control trial (RCT) evaluation. RCT evaluations need to be designed and initiated before the intervention begins. The evaluators randomly assign potential beneficiaries to "treatment" and "control" groups. The comparison of impacts between the two groups reveals the net impact of the intervention. This approach can be used to pilot an intervention to inform programme design. The second main method comprises "quasi-experimental" approaches, whereby statistical techniques are used to construct a control group that can serve as a counterfactual. Such techniques include various forms of matching, instrumental variables designs, regression discontinuity designs, and panel methods. These methods estimate the impact of an intervention by ensuring, to the extent that the data allow, that beneficiaries and non-beneficiaries are similar in all ways except for their treatment status. One such technique is propensity score matching, which estimates the net impact of an intervention by matching beneficiaries with non-beneficiaries who are similar to them in all other ways.

There are obvious drawbacks to these methods for evaluating peacebuilding interventions. Limitations include the frequent absence of baseline data, the desire to "treat" everyone (the humanitarian imperative), the political dangers of random treatment, and the qualitative nature of many relevant variables which are consequently less compatible with statistical methodologies. Impact evaluations are, nevertheless, often possible at the project or programme level and can provide valuable information on what works.

Source: Marie Gaarder (Norad Evaluation Department) and Annette Brown (3iE).

such engagement requires addressing those who have an interest in sustaining the conflict (sometimes called "spoilers" or the "hard to reach").

Questions regarding sustainability might include:

- Which steps have been taken or are planned to create long-term processes, structures, norms and institutions for peace? To what extent has the building of ownership and participation included both men and women?
- Will new institutions designed to address conflict and fragility survive? Are they being used? By whom? Does the intervention contribute to the momentum for peace by encouraging participants and communities to develop their own initiatives?
- Has a meaningful "hand-over" or exit strategy been developed with local partners or actors to enable them to build or continue their own peacebuilding initiatives?

Depending on the activity being evaluated and the scope of the analysis, some specific examples of conflict-related questions on effectiveness include:

- Does the effort result in the creation or reform of political institutions or mechanisms that deal meaningfully with grievances or injustices?

- Have those who benefit from and have a vested interest in on-going violence or instability, or who resist movement towards peace ("spoilers"), been addressed adequately?
- Will improvements in inter-group relationships persist in the face of new challenges and risks?
- Will the parties to a negotiated agreement honour and implement it?

Efficiency

The efficiency criterion is used to assess how economically resources (funds, expertise, time, etc.) are converted to results. In a conflict context, costs associated with prevention work will often be compared with the estimated costs of war or an outbreak of violent hostilities. Averted conflicts, however, are invisible. So there are, unfortunately, not many counterfactuals that compare the efficiency of prevention and conflict – though historical data may be used to reasonably estimate costs.

In addition to comparing the costs of supporting peace with those of war, evaluation could look at priorities in relation to key conflict drivers. It could ask this question: "Is this particular way of working against violence the most efficient option?" In a setting of scarce development resources evaluations should shed light on whether a particular activity is "worth it" compared to other actions or no action. Box 4.6 looks at the security costs of operating in a conflict zone, namely Afghanistan.

Box 4.6. **Security costs**

A joint evaluation of humanitarian aid and reconstruction assistance delivered in Afghanistan between 2001 and 2005 states that there was an overall security overhead of approximately 20% – i.e. operating costs were 20% higher in Afghanistan than in more stable settings for similar types of projects. The evaluation pointed out that these unexpected (or underestimated) costs made the Afghanistan programmes considerably more expensive than similar ones elsewhere. Further evaluative analysis and other data would be needed, however, to determine whether or not they may still be considered efficient.

Source: Chr. Michelsen Institute (2005).

When looking at development or humanitarian interventions in a particular conflict-affected area, evaluators should determine efficiency in comparison to other options for supporting development and peacebuilding in this or similar contexts. Questions on efficiency might include:

- Does the intervention deliver its results in an efficient manner compared to the counterfactual?
- How well are resources (human, financial, organisational) used to achieve results?
- Are there better (more efficient) ways of achieving the objectives?
- What was done to ensure the cost efficiency of the intervention? Did the intervention substitute local initiatives or did it come in addition to local initiatives?

Other criteria: Coherence and co-ordination

Much of the peacebuilding, statebuilding and aid effectiveness literature – and, increasingly, government policies themselves – points to the need for more coherent, better co-ordinated approaches. Funding for a particular conflict prevention project or peacebuilding initiative can be overshadowed and contradicted or – conversely – supported and sustained by other interventions that the same government(s) may undertake. However, the evidence base regarding the use and value of such approaches remains weak. An evaluation's terms of reference may therefore ask evaluators to look at this broader policy context and ask to what extent the activities are coherent with other policies or actions by other parts of government. Evaluations may also look at the degree of co-ordination among or between donors, government and other actors.

Evaluations may assess the costs and benefits of investing in co-ordination and coherence and any unintended consequences arising therefrom. It is important that co-ordination with other actors is not automatically presumed to contribute to better results. Potential risks that have been associated with more co-ordinated donor approaches include: heavy management structures which reduce flexibility and increase overhead costs; strong pressure on national stakeholders as a result of external actors' adopting a block approach ("ganging up" effects); and inappropriate influences on the neutrality, impartiality, and independence of humanitarian and development actors through association with military actors (shrinking humanitarian space). The evaluators must carefully examine the reality of co-ordination mechanisms in relation to the conflict analysis and other criteria (particularly relevance, effectiveness, and efficiency. Bennett *et al.* (2010) discovered in Southern Sudan that, in certain cases, single donor projects were more effective than co-ordinated pooling mechanisms.

Examples of questions that can help assess coherence and co-ordination might include:

- Were coherence and co-ordination factored into inputs and outputs (in other words, were they budgeted for and are they explicitly listed as outputs)?
- What were the roles of and relationships to other actors?
- What aspects of policy were successfully co-ordinated or made more coherent?
  - Joint analyses or understandings of conflict and fragility in the context?
  - Common development of priorities for funding or intervention?
  - Elimination or reduction in duplication of programming?
  - Useful divisions of labour across sectors, issues or problems?
  - Joint evaluations of programming at a country or sectoral level?
  - Useful sharing of experience and generation of lessons?
- Did co-ordination and coherence result in improved effectiveness, efficiency or impacts?
- How much time and what resources were spent on co-ordination? Was it efficient (cost *vs.* benefits) and appropriate? Did it reduce transaction costs?
- How was co-ordination achieved? What were the main constraints and challenges? Is it replicable in other situations (what contextual factors influenced co-ordination and results of coordination efforts)

Draw conclusions and make recommendations

The ultimate goal of any evaluation is to answer the key questions asked and present the results in a useful way. The evaluation report should clearly and credibly demonstrate the link between findings, conclusions and recommendations. Analysis of the collected data constitutes the basis for conclusions. Readers should be able to follow a clear line of evidence that supports the conclusions. Any recommendations should be well founded, relevant, targeted to the intended users of the evaluations, and actionable (meaning those who receive the information are able to act upon it) and should help improve the evaluated activity and any future activities. Evidence-based findings may be a useful basis for discussion between diverse actors in the conflict prevention and peacebuilding fields.

Over time, such conclusions will contribute to better understanding of how peacebuilding and statebuilding processes work, which can help inform strategies for intervention in conflict. Specifically, evaluations can add knowledge by demonstrating the accuracy (or inaccuracy) and use of conflict analyses, or by showing whether or not theories of change work and why. Box 4.7 gives an example of how an evaluation contributed knowledge with findings that yield insight into theories about the links between aid and people's perceptions of aid.

Box 4.7. **Can aid win over local communities? Testing the theory in Afghanistan**

In an evaluation of the impact of development aid on stability in rural communities in North East Afghanistan, the evaluation team examined two theories held by many development actors working in conflict zones. First, the assumption that stability in conflict or post-conflict contexts depends on whether the population perceives the state and its international partners to be legitimate and useful and so co-operates with them, or not. Second, the assumption that aid can influence local people's perceptions of the government and international actors. The evaluation team collected data on the perceptions of the rural population, and matched it with data on aid delivery to the communities. It was found that aid has little impact on how the population perceives international actors. It does, however, lead to more positive perceptions of the state.

Source: Böhnke *et al.* (2010).

Depending on the type of evaluation, conclusions and recommendations may be developed in a participatory format. For example, evaluators could present initial findings to a group of key stakeholders with whom they would then work to draw useful conclusions. On the other hand, if the focus of the evaluation is on accountability, the evaluators are likely to take a more non-participatory approach. Several recommendations from an evaluation of peacebuilding interventions in Southern Sudan are shown in Box 4.8.

Evaluators may discover major differences of opinion not only as to what happened, but as to the value of outcomes and impacts – particularly because individual and group understandings are highly determined by conflict and their own roles in it. Evaluation is, ultimately, a values-based exercise. What is viewed as a successful intervention by some groups may be seen as useless or harmful by others.

Box 4.8. **Making recommendations for donors and the Government of Southern Sudan**

The evaluation team in charge of assessing donor support to conflict prevention and peacebuilding in Southern Sudan (soon to be the independent state of South Sudan) from 2005 to 2010 (Bennett *et al.*, 2010) drew on the evaluation findings to formulate specific, targeted recommendations to donors and the Government of Southern Sudan. Concluding that donors did not adjust what they were doing on the basis of a sophisticated and nuanced analysis of power relations, causes of vulnerability, drivers of conflict and resilience indicators (conflict analysis), the evaluation team recommended that donors and the wider aid community:

- Ensure that revised and new programmes are always preceded by a conflict analysis that links wider dynamics to those specific to the area of operation.
- Plan, monitor and evaluate interventions according to the critical factors identified.
- Rate interventions on responsiveness to conflict factors.

The team analysed which conflict drivers should be prioritised and how this might be done. They suggested the Government of Southern Sudan (and its supporting donors):

- Allocate major resources towards creating and maintaining livelihood programmes for young men who are currently too easily drawn into criminal activity.
- Enable traditional authorities (chiefs) to address root causes of conflict (including disputes over land or bride wealth) at their customary courts by providing capacity-building programmes for these courts.

Develop effective oversight mechanisms to monitor security agencies.

Source: Bennett *et al.* (2010).

Reporting

Reporting takes place throughout the evaluation process (which includes the planning and inception phases). There are often three main reporting steps, though specific reporting requirements may be set by the commissioning agent and should be outlined in the terms of reference.

Inception report

After conducting a conflict analysis and gathering initial information, the evaluators will draft an inception report describing how the team intends to conduct the evaluation and answer the questions set out in the terms of reference. It presents risks and challenges, the methods to be used, data collection tools, indicators if relevant, operationalisation of the main questions, theories of change (implicit or explicit), sample selection tools, case studies (if not selected prior to commissioning the team), the structure of the report, and a work plan for the remaining work. Stakeholders usually comment on the inception reports, often as part of a reference group. (See Chapter 4 for a more detailed discussion of the inception phase.)

Draft report

A draft of the evaluation report is often circulated widely for comments and is a chance for stakeholders to comment on the evaluation. Sufficient time for comments should be built into the overall time frame. Some agencies may also require an "out-brief"

before the team departs from a field visit (between the inception and draft reports) to promote accountability, or other types of reports during field visits.

Final report

Though an evaluation may result in many different outputs, a written report is almost always completed. Reports and presentations will need to be translated into locally relevant language(s) to facilitate sharing with all stakeholders. The final report is sent to stakeholders. Target groups for dissemination should be agreed on at the beginning of the process. Many organisations now use the Internet as an alternative means of publishing the final report, either in part or in its entirety. In all cases, the confidentiality and safety of those who contributed to the evaluation should be carefully protected.

Management response and follow-up action

The recommendations and conclusions of the evaluation should be systematically responded to. The people or institutions targeted (generally donor agencies, implementing agencies or national governments) by each recommendation will respond and take relevant action. The terms of reference should include the process for reporting and responding to the evaluation. A formal response and follow-up by those in charge of a programme will often be required. To write the management response, to disseminate findings and lessons, and to engage in a learning process is the responsibility of the commissioning agency. The form of the management response and follow-up action required varies by institution and may be different for different types of evaluations. A response will often include an assessment of the quality and limitations of the evaluation.

In the case of a joint evaluation, a joint response or multiple individual responses should be used, depending on institutional requirements. In Box 4.8 the Canadian International Development Agency (CIDA) published a formal response to the findings of an evaluation of interventions in what was then Southern Sudan.

Box 4.9. **Canada responds to the evaluation of peacebuilding support in Southern Sudan**

The Multi-donor Evaluation of Support to Conflict Prevention and Peacebuilding Activities in Southern Sudan (Bennett *et al.*, 2010) made recommendations to the Sudanese Government and its international partners, including Canada. Canada has a "whole of government" approach in South Sudan, so relevant departments discussed the report findings. In addition, the Canadian International Development Agency (CIDA) Sudan Programme responded to the evaluation by publishing a formal management response. The response states that CIDA found the evaluation report to be a useful analytical piece. It also pointed out that the evaluation provided a unique learning opportunity because it was the first programme-level multi-donor evaluation in which CIDA had participated.

Box 4.9. **Canada responds to the evaluation of peacebuilding support in Southern Sudan** *(cont.)*

The managers described what they had done or planned to do regarding the specific recommendations. For example, on the recommendation to use better conflict analysis, the managers stated that they "[recognised] the importance of conflict analysis throughout the project lifecycle" and described how they undertook initiatives to promote conflict-sensitive programming, including conflict mapping to guide the design and implementation of new projects relating to children and youth and food security. Specifically, the design of projects together with the United Nations Children's Fund and the Food and Agriculture Organization with the United Nations Industrial Development Organization involved a two-day workshop on a collective mapping exercise to identify potential geographic locations for planned projects and to determine how to address root causes of conflict. Factors related to early recovery needs and opportunities for synergy between food security and youth development programming were taken into consideration.

Source: Canadian International Development Agency (2011).

Disseminate findings

Plans for follow-up and dissemination of lessons learned should be implemented, as agreed during the evaluation planning process described in Chapter 3. Appropriate means of communicating the results to specific target group(s), such as managers, staff and decision makers, should be used. Evaluation managers are increasingly using short summaries, local language translations, policy briefs, video clips and other tailored communication tools to reach different audiences with relevant findings.

Sharing the outcomes of an evaluation can be difficult when the results are perceived as negative or when they question strategies or approaches to which practitioners feel strongly committed. Stakeholders may resist questioning the effectiveness of their approach. Receptivity can be enhanced by emphasising the learning aspects of evaluation – and engaging stakeholders early on.

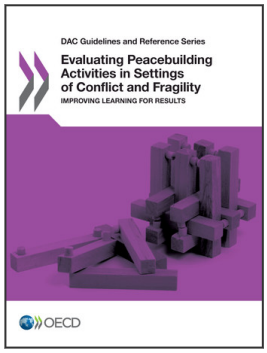
Feed back into programming and engage in learning

A completed evaluation should feed back into the early stages of planning and programme design and help to address challenges by providing more evidence on the validity (or not) of theories of change and data for comparison and reference. Evaluations carried out while a policy or programme is still going on can be used to adjust or redesign it.

In addition to the immediate use of the evaluation findings (by those commissioning the evaluation, for example), opportunities may be identified to feed evidence, lessons or broader conclusions of the evaluation into other policy forums and research activities. For instance, an evaluation of a particular programme working with women peace mediators may yield insights into perceptions of women in different conflict-affected communities which could be valuable to working in the conflict context concerned. Alternatively, it may have lessons for the effective implementation of mediation programmes and contextual factors for success which could be relevant to people working on mediation in quite different contexts.

Evaluators must, in particular, be transparent about the external validity of evaluation findings – i.e. how applicable they are in other contexts. The consequences of applying findings in the wrong way or in a context where they are not relevant can be very negative. This is one reason why it is important for evaluators to explicitly describe weaknesses and strengths of data and methods used, and how they impact findings, conclusions and recommendations.

Having completed the evaluation and learning process, decision makers, managers and staff should be better able to understand and improve strategies, outcomes and impacts, so making more lasting contributions to peace.



From:

Evaluating Peacebuilding Activities in Settings of Conflict and Fragility

Improving Learning for Results

Access the complete publication at:
https://doi.org/10.1787/9789264106802-en

Please cite this chapter as:

OECD (2012), "Conducting an evaluation in situations of conflict and fragility", in *Evaluating Peacebuilding Activities in Settings of Conflict and Fragility: Improving Learning for Results*, OECD Publishing, Paris.

DOI: https://doi.org/10.1787/9789264106802-8-en

This work is published under the responsibility of the Secretary-General of the OECD. The opinions expressed and arguments employed herein do not necessarily reflect the official views of OECD member countries.

This document and any map included herein are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.

You can copy, download or print OECD content for your own use, and you can include excerpts from OECD publications, databases and multimedia products in your own documents, presentations, blogs, websites and teaching materials, provided that suitable acknowledgment of OECD as source and copyright owner is given. All requests for public or commercial use and translation rights should be submitted to rights@oecd.org. Requests for permission to photocopy portions of this material for public or commercial use shall be addressed directly to the Copyright Clearance Center (CCC) at info@copyright.com or the Centre français d'exploitation du droit de copie (CFC) at contact@cfcopies.com.